

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 03:02:45 ; Search time 2798.97 seconds  
(without alignments)  
893.952 Million cell updates/sec

Title: US-09-246-451-1  
Perfect score: 1402  
Sequence: 1 ctgcagatcgttaccgct.....gcgactaccaaagcgtata 1402

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pl2:\*  
11: gb\_pl3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: em\_fun:\*  
17: em\_hum1:\*  
18: em\_hum2:\*  
19: em\_in:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_sy:\*  
29: em\_un:\*  
30: em\_v1:\*  
31: gb\_htg1:\*  
32: gb\_htg2:\*  
33: gb\_in1:\*  
34: gb\_in2:\*  
35: em\_ba1:\*  
36: em\_ba2:\*  
37: em\_hum3:\*  
38: em\_hum4:\*  
39: gb\_pr4:\*  
40: gb\_htg3:\*  
41: gb\_htg4:\*  
42: gb\_htg5:\*  
43: gb\_htg6:\*

44: gb\_htg7:\*  
45: em\_htg1:\*  
46: em\_htg2:\*  
47: em\_htg3:\*  
48: em\_hum5:\*  
49: gb\_pl3:\*  
50: gb\_pr5:\*  
51: gb\_htg8:\*  
52: gb\_htg9:\*  
53: gb\_htg10:\*  
54: gb\_htg11:\*  
55: gb\_htg12:\*  
56: gb\_htg13:\*  
57: gb\_htg14:\*  
58: gb\_in3:\*  
59: gb\_htg15:\*  
60: gb\_htg16:\*  
61: gb\_htg17:\*  
62: em\_htg4:\*  
63: em\_htg5:\*  
64: em\_htg6:\*  
65: em\_htg7:\*  
66: em\_hum6:\*  
67: gb\_htg18:\*  
68: gb\_htg19:\*  
69: gb\_htg20:\*  
70: gb\_htg21:\*  
71: gb\_htg22:\*  
72: gb\_htg23:\*  
73: gb\_htg24:\*  
74: gb\_htg25:\*  
75: gb\_htg26:\*  
76: gb\_htg27:\*  
77: gb\_htg28:\*  
78: gb\_htg29:\*  
79: gb\_htg30:\*  
80: gb\_htg31:\*  
81: gb\_v11:\*  
82: gb\_v12:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1402	100.0	1578	1	PSECCAMC	M12546 Pseudomonas
2	282	20.1	2606	1	PSECCAMRD	D14680 Pseudomonas
3	140.8	10.0	10057	2	AE000101	AE000101 Rhizobium
4	140.8	10.0	34010	1	RSPNGR234	Z68203 Rhizobium s
5	140.8	10.0	236165	5	A79351	A79351 Sequence 2
6	140.8	10.0	236165	5	A93003	A93003 Sequence 2
7	112.6	8.0	3078	1	SERCP450A	M83110 Saccharopol
8	110.6	7.9	5833	2	AF147703	AF147703 Streptomy
9	107.4	7.7	12905	2	AF055922	AF055922 Streptomy
10	104.2	7.4	63734	2	AF127374	AF127374 Streptomy
11	103	7.3	10167	2	SFD08223	U08223 Streptomyce
12	98.2	7.0	2500	2	SS065940	U065940 Streptomyce
13	97.6	7.0	1233	5	E06907	E06907 DNA encodin
14	97.6	7.0	1274	1	STMCYPA50	D30815 Streptomyce
15	92	6.6	1950	1	PSECCAMABA	D00528 Pseudomonas
16	90.4	6.4	7870	1	STE18574	Y18574 Streptomyce
17	89.2	6.4	2218	2	STE250199	AJ250199 Streptomy
18	88.4	6.3	36030	1	MTCY50	Z77137 Mycobacteri
19	88.2	6.3	1629	1	STMSUACB	M32238 S.griseolus
20	88.2	6.3	22574	14	AB018074	Y18556 Cloning vec
21	87.4	6.2	2795	1	AB018074	AB018074 Streptomy
22	87	6.2	1400	1	STWOLEP	L37200 Streptomyce
23	83.4	5.9	1194	5	A48328	A48328 Sequence 6
24	83.4	5.9	4496	5	A48326	A48326 Sequence 4

C	25	82	5.8	39524	1	SCH10	AL049754 Streptomy
C	26	81.6	5.8	6637	2	AF040571	AF040571 Amycolato
C	27	81.6	5.8	107379	1	SHC3PR	X86780 S. hygroscop
C	28	80	5.7	3104	1	STMCHOP4	M31939 Streptomyces
C	29	78.2	5.6	8366	2	AF072709	AF072709 Streptomy
C	30	78	5.6	9882	1	AMOXVAB	Y1652 Amycolatops
	31	77.8	5.5	53784	1	AMM223012	AI223012 Amycolato
	32	77.8	5.5	53785	2	AF69720	AF69720 Sequence 3
	33	77.8	5.5	76199	2	AF040570	AF040570 Amycolato
	34	77.4	5.5	8407	2	SEU82823	U82823 Saccharopol
	35	77.2	5.5	42861	1	WTC1339	Z77163 Mycobacteri
	36	77.2	5.5	107379	1	SHC3PR	X86780 S. hygroscop
	37	76.6	5.5	2168	1	MPOMYCG	D16098 Micromonosp
	38	76.6	5.5	2168	5	E08092	E08092 Genomic DNA
	39	74	5.3	1470	2	AF087022	AF087022 Streptomy
	40	73.6	5.2	1700	1	SGSOYBAC	X63601 S. griseus s
	41	73.6	5.2	1735	5	TI5434	TI5434 Sequence 10
	42	73.6	5.2	4342	2	AF079139	AF079139 Streptomy
	43	73	5.2	47852	2	MTV023	AL022022 Mycobacte
	44	71.8	5.1	2243	3	SEREREFGH	M54983 Saccharopol
	45	71.4	5.1	1688	1	STMSUDCB	M32339 S. griseolus

## ALIGNMENTS

Query Match	100.0%	Score 1402	DB 1	Length 1578
Best Local Similarity	100.0%	Pred. No. 2.9e-283		
Matches 1402	Conservative	0	Mismatches	0
			Indels	Gaps
1	ctgagagatcgtatccgtccgtgcgcgcatctgatatcccaagcgtttctccatgacaagggc	60		
1	CTGAGAGATCGTATCCGTCGCGCATCTGATATCCCAAGCGTTTCTCCATGACAGAGGCC	60		
61	agcaaggaacttgaactggtcgaagcagaagacatgatcaaaacccgtgactccact	120		
61	AGCAAGGCACTTGAACGTGTCAGGAGGAGGACATGATCAAAACCGTGATCGCTCCACT	120		
121	cttaagccaacccgcgtctccagagagaacaacaacttgacagactgaacacataaagca	180		
121	CTTATAGCCCAACCCGCGCTTCCAGAGAGAACAACACATGTGACGACTGAACCATATACAAGCA	180		
181	acgcgaactcttgccccctctgcaaccccaatgtgcgaagaagcctcgtgattccgacttcgaca	240		
181	ACGCGAATCTTGGCCCCCTCTGCCACCCCATGTGTGCCAAGGACCTGTGATTTGCATTCGACAA	240		

RESULT	1	
LOCUS	PSECAMC	1578 bp DNA BCT 26-APR-1993
DEFINITION	Pseudomonas putida camc gene, encoding cytochrome P-450-cam, complete cds and cytochrome P-450-camA gene, 5' end.	
ACCESSION	M12546	
VERSION	M12546.1	GI:151114
KEYWORDS	cytochrome P450cam.	
SOURCE	P. putida (ATCC 17453) DNA, clone PKG300.	
ORGANISM	Pseudomonas putida	
	Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;	
REFERENCE	1 (bases 1 to 1578)	
AUTHORS	Unger, B.P., Gunsalus, I.C. and Sliagar, S.G.	
TITLE	Nucleotide sequence of the pseudomonas putida cytochrome P-450-cam gene and its expression in Escherichia coli	
JOURNAL	J. Biol. Chem. 261, 1158-1163 (1986)	
MEDLINE	86111751	
FEATURES	Location/Qualifiers	
Source	1..1578	
	/organism="Pseudomonas putida"	
	/db_xref="taxon:303"	
	156..1403	
	/note="cytochrome P-450-cam"	
	/codon_start=1	
	/transl_table=11	
	/protein_id="AA25760.1"	
	/db_xref="GI:151115"	
	/translation="MTETIOSNANLAPLPHPVPEHLVFDFDMYPSNLGVOEAMAA VLOSAPYDLVWTRGNGHWMTATRGOLIREAYEYRHSSSPCRIPREAGAVDPIPT SMDDPEOROPALANOVYGCMPVYDKLENRIQELACSLTESLRPGGICNFTEDYAEPFP IRIMWLAGLPEEDIPIHKYITDQMTPRDGSMTFAEKKEALYDILPIIEOROKPCT DAISIVANGQVNGRPRTSDEAKRMCGILLVGGLDTVNNLFSFMEFLAKSPENQELI ERPRIPACQVELLRFSLVADGRILIDSVEFHGVOLKGDQILLPOMLSGLDERENANA CPHWDSRQKVSHTTGGHSHCLTGLQHARREIIVTLKWLIRIPDSIAPQAQIGH KSGIVSGVALPLTWDPATTKAV" 1425..>1578	
	/note="cytochrome P-450-camA (gtg start codon)"	
	/codon_start=1	
	/transl_table=11	
	/protein_id="AA25761.1"	
	/db_xref="GI:151924"	
	/translation="MANANDVVIYGTGLAGVEAFGLRASGEGNINRLVGDAIYIPHH LPLPSKA"	
CDS		
	350 a 490 c 436 g 302 t	
BASE COUNT	5 bp upstream of PstI site.	
ORIGIN		

[illegible]





Db 2565 TGACAAATCCGATGATCTGTGCGCGGTGCAGAGGCCCT 2606

RESULT 3  
AE000101  
LOCUS Rhizobium sp. NGR234 plasmid pNGR234a, section 38 of 46 of the complete plasmid sequence.  
DEFINITION  
AE000101.1  
VERSION AE000101.1  
KEYWORDS  
SOURCE Rhizobium sp. NGR234.  
ORGANISM Rhizobium sp. NGR234.  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.  
REFERENCE 1 (bases 1 to 10057)  
Freiberg, C., Fellay, R., Bairoch, A., Broughton, W.J., Rosenthal, A. and Perret, X.  
TITLE Molecular basis of symbiosis between Rhizobium and legumes  
JOURNAL Nature 387 (6631), 394-401 (1997)  
MEDLINE 97305956  
REFERENCE 2 (bases 1 to 10057)  
Freiberg, C.  
TITLE Direct Submision  
JOURNAL Submitted (22-NOV-1996) Genome Analysis, Institute for Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
REFERENCE 3 (bases 1 to 10057)  
Freiberg, C.  
TITLE Direct Submision  
JOURNAL Submitted (12-DEC-1997) Genome Analysis, Institute for Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
REMARK Update by submitter  
FEATURES  
source location/Qualifiers  
1. 10057  
/organism="Rhizobium sp. NGR234"  
/plasmid="pNGR234a"  
/strain="NGR234"  
/db\_xref="taxon:394"  
<1. 8279  
/organism="Rhizobium sp. NGR234"  
/db\_xref="taxon:394"  
/clone="PXB286"  
complement(206..254)  
/note="nod box-like sequence"  
/note="Region: nod box 13"  
/evidence=not\_experimental  
complement(418..984)  
/gene="y4vD"  
complement(418..984)  
/gene="y4vD"  
/note="hypothetical 21.4 kd protein possibly participating in a redox reaction"  
/codon\_start=1  
/transl\_table=11  
/product="y4vD"  
/protein\_id="AAB91892.1"  
/db\_xref="GI:2182672"  
/translation="MPVKRRVPVAFRTVRDETTGGPNPYRMVEVTTEDYSGRKRV LFSLPGFPTFCSTOOLPDPERLIVDEFKVGIVAVYCLSVNDAFVNNAMKALGEEKV RLPDGGSETRKMGLVAKDNLGFGMSRMRAAYVNDVVEQWFEEDFSDNCEDP VYASSPONTLETRTPDTRARLGRVPIKF"  
misc\_feature complement(1088..1102)  
/note="putative sigma-54 dependent promoter (-12/-24 region)"  
/note="Region: sigma-54 promoter 10"  
/evidence=not\_experimental  
1214..1229  
/note="putative NifA-dependent upstream activator sequence"  
/note="Region: UAS 10"  
/evidence=not\_experimental  
1242..1257  
/note="putative NifA-dependent upstream activator

misc\_feature  
sequence"  
/note="Region: UAS 11"  
/evidence=not\_experimental  
1346..1404  
/note="putative sigma-54 dependent promoter (three possibilities for -12/-24 regions with minimal consensus sequence [GG-N(10)-GC])"  
/note="Region: sigma-54 promoter 11"  
/evidence=not\_experimental  
1442..2134  
/gene="nifQ"  
/gene="nifQ"  
1442..2134  
/gene="nifQ"  
/note="NifQ protein homolog (possibly involved in molybdenum processing)"  
/codon\_start=1  
/transl\_table=11  
/product="NifQ"  
/protein\_id="AAB91893.1"  
/db\_xref="GI:2182673"  
/translation="MSNLAQVRAISKGRVTLGIRMTDRPGMRQLSRLIDGPMPTDL EMDPDQVFAVLSRALREIDAGCATATPATGTSOELDITNRSRPTIVFRLER VVDSERPEEALLRGLLAAHARAGDSASVLPFKITARRALRHDHMLREGLDREVL S RLDASHFTPLAGNTQNMKRRFYRKLCLEAGYSICTAPSCPOCKELESCEGSEGE SASPVGRVGEA"  
2193..3707  
/gene="dctA1"  
2193..3707  
/gene="dctA1"  
/note="C4-dicarboxylate transport protein from the symbiotic plasmid; chromosomal copy exists; significant difference to acc. no. S38912"  
/codon\_start=1  
/transl\_table=11  
/product="dctA1"  
/protein\_id="AAB91894.1"  
/db\_xref="GI:2182674"  
/translation="MRGLRVCMHVEEILIVENLAENVKGRTPHYRHLVQVLAIAV GILGTFYPDVSGKMKPLGDATIMLVKMIIVIPITVATGAGMTDLAKVGRVAGKA MIYFTFESRLALLGLVYVNAVVOVPGAGMIDPASAIDAKIATYAKAHQSOYGLMN IIPITLVGAFASGDILOVIFISVLGICISLAIVKKAENVAVDDLHALTDIFRLVATL M KAPIGAFGAMAFITGKTVASIANLAMLIGTFYLTSLFVPMVLGAARVNGFSIVA LIRYKEELLVLVGTSSSEBALGLNKKMGKCRSVGLVPIGYSFNLDTGNTIYA LTAALFEAQATDTPISYGDQIILLILAMLSKGAAGITGAGFTTLAATSAVSPVA GMAILIGIDRFMSSECRATITNIGNAVATTVVAKMGEELAPAOALTTLGAKAVERMSG LSSQSDVTELEQKVLFGATNSADRTLACRPGCRDSRRATAPHSANQVPGPISL"  
2240..>10057  
/organism="Rhizobium sp. NGR234"  
/db\_xref="taxon:394"  
/clone="PXB368"  
3704..4948  
/gene="y4vG"  
3704..4948  
/gene="y4vG"  
/note="putative cytochrome P-450"  
/codon\_start=1  
/transl\_table=11  
/product="y4vG"  
/protein\_id="AAB91895.1"  
/db\_xref="GI:2182675"  
/translation="MSDLRRKRVKTNIPDPVHPALVHRESLFTSPGNAFPNGOPHA AVACVHDDGPIFTYSPSNTNRDGRGWVITRADQRRVLEDTFSSHRSIFASALGEH MWVPIELDDPRAHGVFRALNLPFSSRRVLALEPTIHAARAGALDICAIEKTSQVMK DPALPFTFVSFLGISOARSEVILVGNVSDLLHGNARRAARSVAVIDENAMAR RKSAPVDEMTFVVOAKTEGRSLTLEEVVRIGVLIVAGIDTVYAAAIIGDMATLANPK HOELLRNPEARGLAAEBELIRAYVQIIRVATKRIEEGCVIRBGDIYSCPAMIANR DTSERCPVITDLARODNCHTAFGTGPHLCHGAHLARREIVIGLEWMLARIPAFRIKE GTAPITTHGGHVIGISNIIITWA"  
4949..5605  
/gene="y4vH"  
4949..5605  
/gene="y4vH"  
/note="hypothetical 24.6 kd protein"

```
/codon_start=-1
/transl_table=11
/product="Y4VH"
/protein_id="AAB91896.1"
/db_xref="GI:2182676"
/translation="MDGRGGSNAPPIAPRSSSLTASHRTMNESTINLFDREAIRDC
LHBYRCRIDADADALRSCYDPARDNHGYSGSAAFIQHALNVEFKRNTIOLN
IILEFLGATEAAVESEFTALORCHDADTEIKOYLCLGRCYCDLEOKREGEMRIYERTVY
YDMWEKORPKSSEAEERFGRPIGAPHDPDPYELSKRRVAKNHFKDILGSGESEP"
gene
5662..7308
/feature="Y4V1"
5662..7308
/feature="Y4V1"
/feature="short chain type dehydrogenase/reductase"
/codon_start=-1
/transl_table=11
/product="Y4V1"
/protein_id="AAB91897.1"
/db_xref="GI:2182677"
/translation="MDROYQGRVIVYVVGAGIGRAIVDFIRANGDVVYVADIPDS
GVIEIGONLCEPHILEVDYSREDDVVALLEKRFKRSIEVLVNNAGICPTAAATAD
TALEDFQRALAINVGAIVSACEFTAKLMKGAALVNVASLAGLGNPKRSAYAAKSG
LISITSLACRMASRGIRVAVAPGHVPTMVALEBRAGLADVSALIRRVPLGRILAP
DEIARAVRFLASQAQSYITGSTLVDDGMMMSVNOPGAGQADORTPAEFMRVEDTD
ARTVIVMGATGAGVGAIALRFAENGDDTVIADGGEEAVKLAGLIDGKHLSRRVDRTV
ETEVVSLFEELRERFGHLDYFVNGMNETIIVPTPESEFVKRLIDVNLGATFCYREA
AISMRSQSVTLNIGASLSLSPFLAPSHAYGNAGIDMTCTCTAAELGPLIRITATVAP
GYIRTCANRLAIVAGMDASLRQRIPLGVNGDAEEVAEAVFLASPDASYINGSTILH
VDGLISSRAGSGSEVDGAISAEMRQRRPARRWLLSP"
gene
7446..8501
/feature="Y4VJ"
7446..8501
/feature="Y4VJ"
/feature="hypothetical 39.2 kd protein; distantly related to
alkanal monooxygenase alpha chains (bacterial
luciferases)"
/codon_start=-1
/transl_table=11
/product="Y4VJ"
/protein_id="AAB91898.1"
/db_xref="GI:2182678"
/translation="MEFAFPIILAOGRVHOSASVIRNSIEQAILSEQAGSTWAFWE
HFNPNYSLVSPILMAHACGLSTIRLGTAVCVLPYQORLAEIGFVANGAL
ELGVSGYQGFEDFRGVNIDEAPFESCDLILGKIKITTHSGRMOIPTAS
VRTLOKTPPIWATATASSKTMARVIRGHNLEPTALHDGETLGLNITKTAAESBG
KEVYDSKVSLLRCYASDDGAEINSYIDNRFQRSEALQQRQOSKDYMEEMPT
HODLSEPTMRKNLPIGSIINRVIDRLLEIDVLKPDOIAIOTOLDPDKTMLNOIELM
GDKIIPAVOKSLRSEA"
misc_feature
8887..8904
/feature="putative NifA-dependent upstream activator
sequence"
/feature="Region: UAS 12"
/evidence="not_experimental"
8996..9010
/feature="putative sigma-54 dependent promoter (-12/-24
region)"
/feature="Region: sigma-54 promoter 12"
/evidence="not_experimental"
Query Match 10.0%; Score 140.8; DB 2; Length 10057;
Best Local Similarity 48.5%; Pred. No. 8,5e-20;
Matches 478; Conservative 0; Mismatches 502; Indels 6; Gaps 3;
```

```
Db 4010 TGGCCGGTCATCCCCCTCGAATTGGACCCCCCGCCCATGAGTGTTCGTGACACTCTTA 4069
Qy 504 aaccaagtgtgttgacagcgtgtgtgtgataagcttggagaacccggatccagaagctgtgcc 563
Db 4070 AATCTCTGTTCTCATCCAGCGGGGTGTGGCTGGAGCGCACTATTCATGCCAGAGCA 4129
Qy 564 tgcgcgtgacagagagcct--gcgccgaagaagacagtgccaacttccagagagactac 620
Db 4130 GCGCGCTGATCGACTGCATTTGCCAAGAGAAAGACCAAGCTTGACGATGAAGATTTTC 4189
Qy 621 gccgaacccctccagatacgcatactcatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 680
Db 4190 GCTTGGCCCTTTACGTTTCAGCGTTTTCCT-CAGCTTTCGTGGAGCTTTCGCCAGAGCGCATC 4248
Qy 681 ccgcacttgaataactcctaagaagatcagatgaacccgtccagatgagcagatgaccttgcga 740
Db 4249 CGAAGTACTTGTGGCTGGGTGAACGATTTGCTCCACGCGCAACGCAAGCAAGG-- 4306
Qy 741 gagcccaagagagcgctcttaagatactgatacgcatacgcagagcaagcgagcgagaag 800
Db 4307 GCACAGCGCCGCTGCTGCTGCTGCTTTATTTGACGAATGCGAGCGATCCGCAAGTCG 4366
Qy 801 ccgggaaccgacgctatcaagatcgtgtgccaagcgccaggtcaatggtggtgacgacacc 860
Db 4367 CCAGCTGTGCATTTTCATGACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4426
Qy 861 agtgaagaagcgaagaagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 920
Db 4427 GAAGAGGAGTCCCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4486
Qy 921 aattccctcagctcagatgagtgagttccctgtgccaagaagcccgagagcgtcgcagaagctg 980
Db 4487 GCCGCCATTTGCTTTGACATGCGCTATCTTTCACGCAATCCAAAGCACAGAGAGTTGTTG 4546
Qy 981 atcgaagcgtcccgagcgtatccagcgcgtgtgaggaagaaactacccgtggtgtgtgtgtgt 1040
Db 4547 CGAAGCGAACCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 4606
Qy 1041 gttccagatggtcgcacatccatccatccatccatccatccatccatccatccatccatccatcc 1100
Db 4607 GTTCAGATTAATCCCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 4666
Qy 1101 gttacacagatccctgtacagcagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1160
Db 4667 GGTATTTATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4726
Qy 1161 ccgattgcagcgtcagatcagtcgccaagaagtttccacacacacacacacacacacacacacacac 1220
Db 4727 CCCAACAATATCGATCTGCGACAGACAGATTAACACACACACACACACACACACACACACAC 4786
Qy 1221 catctgtgcttggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1280
Db 4787 CACCTTTGCCACGAGGCGCATCTGCGACGGGAGAAATCGTCATTTGGCTACGTGAATAGG 4846
Qy 1281 ctgacccagatctctgacttccat 1306
Db 4847 TTAGCGCGCATCCAGCTTTCGCAAT 4872
```

RESULT 4  
LOCUS RSPNGR234 34010 bp DNA BCT 08-AUG-1996  
DEFINITION Rhizobium sp. plasmid NGR234a DNA.  
ACCESSION 268203  
VERSION 268203.1 GI:1486419  
KEYWORDS BioA: C4-dicarboxylate transporter; CapA: cyrP450; FabG: rdxN;  
FixA: fixB; fixC: fixX; fixY: gmpD; glutamate dehydrogenase: hmdA;  
LuxA: NifA; NifB; NifD; OppB; OppC; oppD; OppP; transposase.  
Rhizobium sp.  
Rhizobium sp.  
Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group;  
Rhizobiaceae: Rhizobium.

REFERENCE 1 (bases 1 to 34010)  
AUTHORS Freiberger,C., Perret,X., Broughton,W.J. and Rosenthal,A.  
TITLE Sequencing the 500-kb GC-rich symbiotic replicon of *Rhizobium* sp.  
NCR234 using dye terminators and a thermostable 'sequenase': a  
beginning  
JOURNAL Genome Res. 6 (7), 590-600 (1996)  
MEDLINE 96389014  
REFERENCE 2 (bases 1 to 34010)  
AUTHORS Freiberger,C.  
TITLE Direct Submission  
JOURNAL Submitted (13-DEC-1995) Christoph Freiberger, Genome Analysis,  
Institute of Molecular, Biotechnology, Beutenbergstrasse 11, Jena,  
07745, Germany  
FEATURES  
source Location/Qualifiers  
1..34010  
/organism="Rhizobium sp."  
/plasmid="pNCR234a"  
/strain="NCR234"  
/db\_xref="taxon:391"  
/clone="cosmid PXB296"  
/clone\_1lb="genomic cosmid library"  
1..625  
/gene="orf1"  
<1..625  
/gene="orf1"  
/note="putative"  
/codon\_start=2  
/transl\_table=11  
/product="OppB homologue"  
/protein\_id="CA92398.1"  
/db\_xref="GI:1486420"  
/db\_xref="SWISS-PROT:Q53191"  
/translation="ILSVYVSGFVGLAAMRSGGFVDRALATFSAIGSVPEVIGYI  
LTFEALIQTRMLPVGGTYSINCGVAPPELHLILPTVTLISVPTAFIARITRSGMLEVL  
SEDMYRAAKGKSPFAMLFHMLKNAQVPIILVIGISFPMIGGVLTFTYVNWGCI  
GRLVDAIKRDPILTIQTVLVLSGLVILNLVDLAVTLIDPRIRI"  
628..1503  
/gene="orf2"  
628..1503  
/gene="orf2"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="OppC homologue"  
/protein\_id="CA92399.1"  
/db\_xref="GI:1486421"  
/db\_xref="SWISS-PROT:Q53192"  
/translation="MALTTSRASGSPRIQTHVYRIAKRHPLVLLGGTILLILLAL  
AAPLYSGDPLMDPFRLOQPSASMMFGTDNLGRVFARTIYGARISLIYGLLSAVCA  
AVCGLLIGVLAGSRTEPNDIMKRVMDGLMSIPFLALIALSLTGPQIGLLIYAIAP  
EPFAYTRLVRSVLYSVSRPYEALCGGARLPVLRHILPSTIPPLMWQATVCAS  
AIMEAGLSIFIGVSPSEIPSGWGNMANSRLFLAIALPTIFAPGLCAVTVLVANLIG  
DGLRDMFDPRSKRRR"  
1505..2512  
/gene="orf3"  
1505..2512  
/gene="orf3"  
/note="putative: at pos. 1443 also T possible without  
changing the encoded amino acid"  
/codon\_start=1  
/transl\_table=11  
/product="Oppd homologue"  
/protein\_id="CA92400.1"  
/db\_xref="GI:1486422"  
/db\_xref="SWISS-PROT:Q53193"  
/translation="WRKQSNPKADLLLDVRLDLETHYGYEESVTALGCYNFHVKSGEM  
LGIVSGSGCKSVTALSLRLPLKQTAIKVSGVITKGRMLLELSRQMEVGNINIA  
MIDQDPMTLSPIVHTVGRQIAEAVQIHGRVSALEKALEMLRLVRIADPERLLANY  
PHEMSGMRORAIAMALACSPELLIADDEPTLADVTIOAOIQLRLIYDLDRMGTAVM  
FIYHDGVVAETGCRQVITYAGRIYVQASVTDLFARPTHTYTRALNNSVDRRGSGS  
RIPEIRGVVPSLREPLVCGSFARCEPFAIGVCERKMPVLSEIQPGHAACWRNNEVYN  
L"  
2509..3570  
gene

CDS  
/gene="orf4"  
2509..3570  
/gene="orf4"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="Oppf homologue"  
/protein\_id="CA92401.1"  
/db\_xref="GI:1486423"  
/db\_xref="SWISS-PROT:Q53194"  
/translation="MSDALIKVESLTKKHPITENGWFKTNPVYRAVEDSVFTVQAGET  
LCVYSGSGCKSTIALKRLKLDIPISGRVLYESTDILAGLKNKELRAMRRMOMVQEDP  
YSSLNRLTAGQIITPEVNEFCLSRQKQELATDLRKVGMSPEMASRYPSMSGQ  
RQRIQIARALALKPSLIIADENAVSALDVQAOIIMLMDLOETGIALIFSHDLAY  
VEHIGRVAVMYIGRIVELSPRDALCPVHPYTEALIAAPVDPDARARLEVPLEGE  
VPSPVNPSSGCAHPRCPPLAVDRCAVEPPLVPMQGRVACHVRAPAAAHISBAL  
ASLALPKTYQPGGTILARR"  
3606..4991  
/gene="orf5"  
3606..4991  
/gene="orf5"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="CapA homologue"  
/protein\_id="CA92402.1"  
/db\_xref="GI:1486424"  
/db\_xref="SWISS-PROT:Q53195"  
/translation="MERCSESGRHSABENDKRTIDTGSTATNVVDGFTVAVGADVTS  
RALANGHHPGSEIVELLRADVTFCNMETLLPDIRSFNGTQDAEGVAGVHSLPEIG  
PDLKAMGFNIMGRRANNHSLDMCVGEGRETSRIIDESGIIHAGVGEASAASARLIET  
ARGRVALLSCATSEFTPMRSRACDPAGCAPARPGVNALRLRSVVPEPMDASRLIKIDA  
LNPQGHKHDREMLVLAGTTRGKGVGYTAANTRLADILRNVRGKQYSDPCTF  
NNAHPEGMSSEEPADDEQALRLIDAGADAYVGHGPHRLRGIEIYKRPDIYSGCNF  
PDDLTETPGADMTVDYDKDPQVDTPAEPTAAEETGTPTAAGFTGALKEPYYESV  
AVSRFENQIAELRLPIELGYSKRLANGVPSLAPRPAISILERIQRLSBPEGTRI  
TIEDRVGLRL"  
5460..6863  
/gene="orf6"  
5460..6863  
/gene="orf6"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="BioA homologue"  
/protein\_id="CA92403.1"  
/db\_xref="GI:1486425"  
/db\_xref="SWISS-PROT:Q53196"  
/translation="MTIDIKDISKDRNTVLHPFQOLKDPATGKLRPTIVETGKGR  
IODARNGNOLIDGPAGLYCVNAGVGEAEVAEISRQYRLAYHSYAAHTDELAISLD  
RLVKMAPGKMSKVEYCGMSDANETQAKLWYVYNLRGKPTKKIISREGYHGCSTV  
SGSMGMSRYHOMDPLPOLICHTGYPHHYKCANCPETREESARPAALDPMETIG  
PNVNGAFIAEPVLGTGCTPPPEGTVEAIOAYLKKHVDLLIDEVITGGRGSMFGS  
OHYGEIPDLITYAKGLTSAVFPPLSASIVEKVKYKVLDEGDADRVGASHGVTYSGHPIG  
AAANAVNIDIVEKEKEDPGNAREVGYVQALKEKFNQDPIVEGRVGVLMGAIIEFVGD  
RENNKRFDEPLKVGARVSKAARDRGILIARAMPHGDILGAPPLVTYTKEEVDEIVAAE  
KAVRSMDELVDRDQKL"  
6888..8426  
/gene="orf7"  
6888..8426  
/gene="orf7"  
/note="putative: GNG start codon; alternative start  
codons: 6912; 6927; 7017"  
/codon\_start=1  
/transl\_table=11  
/product="GapD homologue"  
/protein\_id="CA92404.1"  
/db\_xref="GI:1486426"  
/db\_xref="SWISS-PROT:Q53197"  
/translation="MDAGRRRLVINDCMPEGRGRSPILPRPRTPPDESSGIGINAM  
TTRAKLMTTSLAPIAVHSPPYDGSLLASVEATPDADIDRLATARCAETSRNLPKRR  
ASTLEGAQOVESRDHAEIIVREAGKTIYQARKVELCVMTLAKISAEAAKRNAGEI  
VPFDAYTGEORGWETRPDLGIITPAITFYNPLNLVNAHKLGPAIAGNAVMILKPSNL

[illegible]

/note="putative; alternative start codon: 11220"  
 /codon\_start=1  
 /transl\_table=11  
 /product="glutamate dehydrogenase homologue"  
 /protein\_id="CA992406.1"  
 /db\_xref="GI:1486428"  
 /db\_xref="SWISS-PROT:Q53199"  
 /translation="MSFLDIPSLPRRIIIQCSNPTVTFVGLRGRMYSFGMSVRE  
 HCEYKQDIPASNAEAEVEALALATKLSLVPPGSGSGALIKDIPRMSQFQELE  
 IHTFRFQENMKRPDARARQCVASDITGTERMMAMMEFRFANTVDVSGACVTF  
 PLSGSGADGAASTGREGVQVANSI,SRDTRFDGLGRNLINAGSVTIGFEGNVYAA  
 RLSESGDAVTVTAERDGVYANPELSTLEAKKHQITGSGIDLPDQAKSIAQMGCV  
 EOPCDVLIPIAMERAIHAEBAEKMKHLVVEAKNPQVFEADEILRSQGVITLDDLV  
 NAGGVVYSYERKKNITITPEGLFEMRRRREKRNITATLRMGKESPSAMPDRQEFLE  
 TCAIDVIRSGLEQVEMWSTVTRADLMEQOPELGTRYAAVVAAYRQAD"

[illegible]

Db	29976	CGAGTACTTGTGGCTGGTAAGCGATTTCCTCCACGGCAACGCAGAAACCGAAGS--	30033
QY	741	gaagcgaagagagcgctctacgactatctgaaccatcatcagcgaacgacgacagaag	800
Db	30034	GCAGCAGCCCGCTCGTGGCCCTTTATGACGAAATAGGACGCGATCGCCGCAGATGC	30093
QY	801	ccggaagaccgacgcatacgaatcgcttcgcaacgcccagctcaatggtggcgacatacc	860
Db	30094	CCAGTGTGTCATTATCATACCTTCGTCTCTCAGCGGAAATCGAGGGCGCTCTTAACA	30153
QY	861	agtagcgaagcgaagagatgtgtgacctgtactcgtgcgacgacctgatacagtgctc	920
Db	30154	GAAAGAGAAAGTCCCGGCGATGGGTGCTTTTCTTGTGCGGGGGCTGGAACCGTTGCC	30213
QY	921	aatttcctcagcttaagcaatggagtctctctgaccaaagcccgagacatcgccagaagctg	980
Db	30214	GCCGCCATTGGCTTTTGAACATGGCCTATCTTTCACGCGCAATCCAAAGCACACGAGACTTGTG	30273
QY	981	atcgagcgctcccgagcgatctccagcgccttgcgagaaactactccgacgtcttcgctg	1040
Db	30274	CGAAACGAAACCGGCTCGGCTGGGCTGGCGCTGCCTGAGGAACGTGCGGCTATTTCAMC	30333
QY	1041	gttcgcgattgcccgcacatccctaactcgcgatacgaagttcaatgctgacgtgcaatgaagaa	1100
Db	30334	GTTCGATTAATCCCGTGGCGCACGAAAGACATCGAGTCGAAAGCGTCCCTATCCGTAG	30393
QY	1101	ggtgacacgaatccctgctacgcgaatgtgtgtgcctgtagtgcgagacgaaacgcctgc	1160
Db	30394	GGTGAATTATGTTTCTCTGTCGCCGCGATATTGCTAATCGAATCCGCTCTGAATTTAAGTGC	30453
QY	1161	ccgattgacgctgcgacttaagtcgcgcaaaagttcacacacacaccttggccacgagac	1220
Db	30454	CCCAACACTATTCGATCTGGCAGCAGAGAGATTAACCGACACACCGCTTTGGCTATGAGACA	30513
QY	1221	catgtgtgccttggcgcgaacaccttgcgcgcgagaaatcatctgtcaaccttaagaatgt	1280
Db	30514	CACCTTTGCCAAGGAGCCCATCTCGCACGGCAGAGAAATGTCATTTGGCTACGTGAAATGG	30573
QY	1281	ctgaccaggaatcctgacttctcaat1306	
Db	30574	TTAGCGGCATCCACGTTTTCGCAT30599	

QY	324	actcgcctgcacacgagcgacatgcagtcgcacatccgcgcgcacactgcctcgttcagccctat	383
Query Match	10.0%	Score 140.8; DB: 5;	Length 236165;
Best Local Similarity	48.5%	Pred. NO. 6.2e-20;	
Matches 478;	Conservative	0; Mismatches 502;	Indels 6; Gaps 3;



Oy	1101	ggtgacgcgaatccgcgtctccgcagatgctgtctggtccgtgtagtcagcgcgaagcctgc	1160
Db	147562	GGTGGATTATGTTTCTCTGTCGCGCGAGTAGTATCTAATCAGATCCGTCGTGAAATTTAAAGTC	147621
Oy	1161	cgaagcagcgtcgaactcagtcgcgcgaagaagtttcacacacaccttggccacgcagc	1220
Db	147622	CCCAACACTATTCGATCTGTGGCAGACAGATTAACACGACACACCGCTTTGGCTATGACCA	147681
Oy	1221	catctgtgccttggcagcagccttggccgcgcgggaatcatcgtcaacctcaagaaatgg	1280
Db	147682	CACCTTTCCACAGGCGACACTCTGCACGCGGAGAAATGCTCATGTCGCTACGTGAATGG	147741
Oy	1281	ctgacccagatctcctgaattccat	1306
Db	147742	TTAGCGCGATCCACGCTTCCGAT	147767
RESULT	7		
LOCUS	SEKCP450A	3078 bp	DNA
DEFINITION	Saccharopolyspora erythraea ORF 1 gene, partial cds; cytochrome	ECT	26-APR-1993
ACCESSION	P-450 gene, complete cds; ORF 2 gene, partial cds.		
VERSION	M83110		
KEYWORDS	M83110.1 GI:152682		
SOURCE	cytochrome P450.		
ORGANISM	Saccharopolyspora erythraea (library: NRRL 2338) DNA.		
REFERENCE	Saccharopolyspora erythraea		
AUTHORS	Bacteria; Filicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.		
TITLE	1 (bases 1 to 3078)		
JOURNAL	Andersen, J.F. and Hutchinson, C.R.		
MEDLINE	Characterization of Saccharopolyspora erythraea cytochrome P-450		
FEATURES	genes and enzymes, including 6-deoxyerythronolide B hydroxylase		
SOURCE	J. Bacteriol. 174, 725-735 (1992)		
	92121109		
	location/Qualifiers		
	1..3078		
	/organism="Saccharopolyspora erythraea"		
	/db_xref="taxon:1836"		
	/tissue_lib="NRRL 2338"		
	<1..859		
	/note="ORF homologous to Escherichia coli and Salmonella		
	typhimurium dnaE gene; putative"		
	/codon_start=2		
	/transl_table=1		
	/protein_id="AA26482.1"		
	/db_xref="GI:152683"		
	/translation="AGCGRGRAQAPAGPARVPVRRRRRRGVVAAAPAVHAG		
	RVAQAQALLRACGAPLRGPSPGRRAAARVPGRHRRARRRRAGGPGQDRR		
	DDCDPAAPHOERPPVGDPRDGRRCRGAVLPEVLADVRLLPARGHRDRGEPHQR		
	ARGHQLRLRRAGGHLGGRDRPGRNPVRDQAGGQGVGAKRAQHLAQLGIG		
	PGAEAAAGAARGHPRAVVRVRLVRDERRAAGRAAGVLRDGAASRGAASPLRA		
	PRPRLRGS"		
	949..2166		
	/codon_start=1		
	/transl_table=1		
	/product="cytochrome P-450"		
	/protein_id="AA26483.1"		
	/db_xref="GI:152684"		
	/translation="MTTGEVPDLADPDAFAODRHNRYARREPVORIRVINGLDAM		
	LITPEYVKOALLDPRIAKDGFRPOOITIERIADARRPGSPDIGHMINTPDDHT		
	RRLKLVYKATFARVEGLRRRIEDITDDLLDRLAGREVDLIDFAPPLPTIYSELM		
	GVESRRDDFRSMTNVLVDGSQPEAQAQSAVAMVEYLTLEIAKRTPEPGDLSLALLE		
	AVEGDRLLSGELLAMVLLDVAGHETTVNLIGCVSLTGLNPDLAALRNDSPTLALGE		
	AIETETLYRESVPANGTEPRHTAEAVRFCDVVIPEELTWLVLGANRGERFEEDRFD		
	ITRETTGHAFAHGCHIFCCVGAALARTLBAQIAVAGRLRLRPDLURMAASPDLLRMFVTL		
	MRCLEXLPVRPGA"		
	complement(2365..>3078)		
	/note="ORF 2"		
	/codon_start=1		
	/transl_table=1		

Query Match	Best Local Similarity	Score	DB 1	Length
Matches 410; Conservative 0; Mismatches 469; Indels 6; Gaps 1;	46.3%;	112.6;	DB 1:	3078;
468	galtccgcgaagcaacgagcttctgctgcgcgcgcgaacaaacagtggttggcagtcgcggtg	527		
1237	GACCCGCCGACACACACCCGCGTGGCGCAAGCTGTGTCTCAAGCCCTTCAACCCGCGCGCG	1296		
528	gtggaataagctggaagaaaccggaatccagaagctgacctgtctgtatctgagagacctgcgc	587		
1297	gtcgaagggcctgcggcccgcgatcagacagatccacgacacgtctgcacccgctgcgcg	1356		
588	ccgcgaagacagctgaacttcaactcagagagactcgcgcgaaccccttcgcgaatacgaatcttc	647		
1357	ggccggtccgaagtgatctgattcgaacagcgatggtccggttcccgctccatccatccatgattc	1416		
648	atgctgctgcgaagctctaccggaagaagatattccgcgaacttgaaatlaactaagcgaataag	707		
1417	tccgagactgattggcgctgagagacttccgcccgcacacatctccgctccgtccgacacacgctc	1476		
708	atgaccgcctcgcgatgacgacatgacatcttcgcgagagcgcaagagcgctctacgaactat	767		
1477	ctggtctcgaaggggtccagacgcccagagccagccagcccgctgctggtgctgagatc	1536		
768	ctgataccgatactatgagcaacgacgagcagaagccgggaaaccgacgctatacgaatcgtc	827		
1537	ctgaccgacgtgatggtcccaaaagggcaccacgacccggcgacgacgtctgaccccgctg	1596		
828	gccaacgagcgaatcgaatgagcgacacagatgacagtgacgaagccagaagagatggtgtgcg	887		
1597	ctggaagcgcgtgagagacgacgacccgctctacacgagggcgacgtatgccattggtgttc	1656		
888	ctgttaactgctgcgagcctgtgatacggatggttcaattctctcaagcttcaagcagtgatc	947		
1657	ctgtgctgtgtgtgcgcgggacagagacggtcaacctgattcggcgaactcgcgtgctgcg	1716		
948	ctggccaaaagcccgagacatcgcaagagatctgatcgagcgtcccgagcgtattccagac	1007		
1717	ctgctggtggcaaccccgacacagcttcccgcgccgacacgacccctctgctgctccgac	1776		
1008	gcttgcgaagaaactaactccgcgcgttctcgtgtgttgcagatgagcgcaatccctcaactc	1067		
1777	ggcattcgaagaaaccttgcgcttaccagagaccccggtggtggcaagggacgcttccggacac	1836		
1068	gattacgaattcattgycgttgcacttgaaga-----aagtgacacgaataccctgtatcag	1121		
1837	gccgaagcgcgctccgggttgcgcgacgctggtgattcccgagagagacgtctgctgggtccg	1896		
1122	cagatgctgcttgcgcttgaatgagcgcgaaacgccttgcgcgaatgacgctgacttcaagt	1181		
1897	ctcggggcgcggaacccggagcggtagagcgttgcgagagaccccgacccgcttgcacatcacc	1956		
1182	ggccaaaaggtttcaacaacacacacttcttggcagcgagacacatctgtccttggcagaac	1241		
1957	cgggaagaccacccgacgcttccgttggccacacgacatccacttctgctggcgcccgcc	2016		
1242	ctgcgcgcgcggaataatcgttaacccctcaagaaatgctctaacgaagattcttcaactc	1301		
2017	ctggcccgcgctggaagccacagatccgctgacagcgtcttctgagccgcttcccgagactg	2076		
1302	tcatttgcgcggtgtgcacgaattcagacaagaagcggtacgtc	1346		



Db 2077 CGGATGGCCGCGGAGACCTGCGGCGGTTTCAGCGTC 2121

RESULT 8  
AF147703/c  
LOCUS  
DEFINITION  
AF147703 5833 bp DNA BCT 01-JUL-1999  
Streptomyces fradiae demethylmacrocin O-methyltransferase Ty1E  
(Ty1E), hypothetical NDP-hexose 4-ketoreductase Ty1D (Ty1D),  
tylodoxin (Ty1H), cytochrome P450 Ty1HI (Ty1HI),  
macrocin-O-methyltransferase Ty1F (Ty1F), and putative NDP-hexose  
3-epimerase Ty1J (Ty1J) genes, complete cds.  
AF147703 GI:5305783

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
SOURCE

Streptomyces fradiae.  
Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 5833)  
Bate,N.,. and Cundliffe,E.  
The mycinose-biosynthetic genes of Streptomyces fradiae, producer  
of tylosin  
Unpublished  
2 (bases 1 to 5833)  
Bate,N.,. and Cundliffe,E.  
Direct Submission  
Submitted (30-APR-1999) Biochemistry, University of Leicester,  
University Road, Leicester LE1 7RH, UK  
Location/Qualifiers  
1..5833  
/organism="Streptomyces fradiae"  
/strain="J59235"  
/db\_xref="taxon:1906"  
complement(72..1259)  
/gene="ty1E"  
/note="orf24\*"  
complement(72..1259)  
/gene="ty1E"  
/function="converts demethylmacrocin to macrocin during  
tylosin production"  
/note="2'-'-O-methyltransferase"  
/codon\_start=1  
/transl\_table=11  
/product="demethylmacrocin O-methyltransferase Ty1E"  
/protein\_id="A041815.1"  
/db\_xref="GI:5305784"  
/translation="MAVOKKATLVKQIIRAAGSHADREIVAEHGPVATVYDEI  
VSRAPVADVPLVELAVRSGDALVPRRLAQAQAPVRAAPDDGCVAMRYEELD  
ELVRELFGPCREKAAAGTGTTLFPVATSGHIDYFLAAQATATATVLAGCTSAKPD  
NELVTSYLPKMGSLHMFPHYDRHREYRNEVRVLEIGIGQHPHMGGSGLRMWK  
HEFHRLITGLDIEDKSHAEORITTVGNDPNDPGITTELARVGFEDYIDGSHIN  
EHVRSFHALFPHVRPGRGLYVIEDLWTAWPVGGSGDPGKSDLTSLGIVKSLVDSLO  
HQLPDSGSRPCYADRHVYGLVHNLAFIEKGVSEGGIRGMITRDPDALVAASG  
GA"

gene  
complement(1277..2287)  
/gene="ty1D"  
/note="orf23\*"  
complement(1277..2287)  
/gene="ty1D"  
/function="involved in synthesis of NDP-deoxyallase during  
tylosin production"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical NDP-hexose 4-ketoreductase Ty1D"  
/protein\_id="A041816.1"  
/db\_xref="GI:5305785"  
/translation="MRAAGDWAGRVVVTGALGIGSHYAEKIALGARGVGLHNSAR  
PAAGAPAAPAAGRSCLVADLCDEAATRADEPAGADVTHCAGLDGNADYKRDH  
SASVLDANVTAGTAVNTADTGAGAVLLSDEVYCAPDSPAEDDEIRVRYPHAG  
NGVLSKICEIMAEELHGAEEGSRIFRVGNYGPRDGGGRTRTVPSMVAARA  
ETLEWGDGROFRSFTHVEDLVNATLRMEADKPYPMNIGSPREYSILDLARTVSAV  
GKGGRITRYPDKSPPAORLDSNRKDVTCGEPRLVRGIGIOTITMYRKHHPSPAP  
HS"

gene  
complement(2284..2529)  
/gene="ty1HI"  
/note="orf22\*"  
complement(2284..2529)  
/gene="ty1HI"  
/note="ferredoxin; Ty1HI"  
/codon\_start=1  
/transl\_table=11  
/product="tylodoxin"  
/protein\_id="A041817.1"  
/db\_xref="GI:5305786"  
/translation="MRVRIIDTGCVCAGAGCCERAPPTVFRODEGVGVLDKTPPAW  
EEVREAEDLCPPARAVALSGDGTGAGAGAAPPTGRDA"  
complement(2565..3875)  
/gene="ty1HI"  
/note="orf21\*"  
complement(2565..3875)  
/gene="ty1HI"  
/function="hydroxylation of polyketide ring at C-23 during  
tylosin production"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P450 Ty1HI"  
/protein\_id="A041818.1"  
/db\_xref="GI:5305787"  
/translation="MAMADPTVFSLRPQVSSSGDAPRSOKGILLPAARANDTDEA  
GRSITAMPVAPRCPPSPPEQVAAALRAEPIARALMDGAPVWLISRODHVRLADPR  
VSHHPAKLRLSPSDEAFAASRLTLPDPDHGALRGHPIPERGLARVELRPSVQI  
VTGLDDLTARGDADLADFALPMATQVIGRLDIPYEDKDTFQERTQATRPAGE  
EALFALLELDVYLDRLISGKTGREGSDGMVAVARCGGSHADVLNVALLAAG  
HETTSMTVMVSVLYVLQHPPTAMRELTVNGLLPAGVDELRLYLSIDGLRSATADIE  
IGCHTIRAGDGLVFLAANRDEAVFSEEPADIIHRSARRHVAFCYGPQCIGQNLAR  
MELEVALGAVLEELPALRPTTQVAGLRKSDSAVFVEYELPVAM"  
4090..4860  
/gene="ty1F"  
/note="orf20\*"  
4090..4860  
/gene="ty1F"  
/function="converts macrocin to tylosin"  
/note="3'-'-O-methyltransferase"  
/codon\_start=1  
/transl\_table=11  
/product="macrocin O-methyltransferase Ty1F"  
/protein\_id="A041819.1"  
/db\_xref="GI:5305788"  
/translation="MABSPDHARDLYIELKKVSVNVIEDPPTHVAGMTDASFDRTS  
RESGEYPRVAHMTGLKRLDNLHRLAVVEDGVGDIETGVWRGACITARKGLN  
ATGQADRTVWADSFQGFPELTGSDHPLVEIDLHOYNEAVDLPSTSETVRENFARYG  
LDDNVRFLAGWKFDTMPAPVAKQALVMRLDGDYGTATMDVLSILYERLSPGVYIVD  
DVCIIPACREAVHDFRDLGIRPTIHRIDQAGVWRHSG"  
4897..5514  
/gene="ty1J"  
/note="orf19\*"  
4897..5514  
/gene="ty1J"  
/function="synthesis of NDP-deoxyallase during tylosin  
production"  
/codon\_start=1  
/transl\_table=11  
/product="putative NDP-hexose 3-epimerase Ty1J"  
/protein\_id="A041820.1"  
/db\_xref="GI:5305789"  
/translation="MRDTRRPLGIEGAVYIOPEIHPDRGEFFHAMFOSSEPRRTGHS  
FSPYOVNIAVSRKGLRGTHFSVPVPGAKYSACVAGAVEVYVDRLGSPYGGWRA  
VPDEYNRTAVVYPACLGAFVALTDRTTLVYLCSESYARREHAYNPDLPLATAMP  
EDVELLSRDRAPTLAEARKGILPSTQETREHGLPEHARR"

gene  
complement(2284..2529)  
/gene="ty1HI"  
/note="orf22\*"  
complement(2284..2529)  
/gene="ty1HI"  
/note="ferredoxin; Ty1HI"  
/codon\_start=1  
/transl\_table=11  
/product="tylodoxin"  
/protein\_id="A041817.1"  
/db\_xref="GI:5305786"  
/translation="MRVRIIDTGCVCAGAGCCERAPPTVFRODEGVGVLDKTPPAW  
EEVREAEDLCPPARAVALSGDGTGAGAGAAPPTGRDA"  
complement(2565..3875)  
/gene="ty1HI"  
/note="orf21\*"  
complement(2565..3875)  
/gene="ty1HI"  
/function="hydroxylation of polyketide ring at C-23 during  
tylosin production"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P450 Ty1HI"  
/protein\_id="A041818.1"  
/db\_xref="GI:5305787"  
/translation="MAMADPTVFSLRPQVSSSGDAPRSOKGILLPAARANDTDEA  
GRSITAMPVAPRCPPSPPEQVAAALRAEPIARALMDGAPVWLISRODHVRLADPR  
VSHHPAKLRLSPSDEAFAASRLTLPDPDHGALRGHPIPERGLARVELRPSVQI  
VTGLDDLTARGDADLADFALPMATQVIGRLDIPYEDKDTFQERTQATRPAGE  
EALFALLELDVYLDRLISGKTGREGSDGMVAVARCGGSHADVLNVALLAAG  
HETTSMTVMVSVLYVLQHPPTAMRELTVNGLLPAGVDELRLYLSIDGLRSATADIE  
IGCHTIRAGDGLVFLAANRDEAVFSEEPADIIHRSARRHVAFCYGPQCIGQNLAR  
MELEVALGAVLEELPALRPTTQVAGLRKSDSAVFVEYELPVAM"  
4090..4860  
/gene="ty1F"  
/note="orf20\*"  
4090..4860  
/gene="ty1F"  
/function="converts macrocin to tylosin"  
/note="3'-'-O-methyltransferase"  
/codon\_start=1  
/transl\_table=11  
/product="macrocin O-methyltransferase Ty1F"  
/protein\_id="A041819.1"  
/db\_xref="GI:5305788"  
/translation="MABSPDHARDLYIELKKVSVNVIEDPPTHVAGMTDASFDRTS  
RESGEYPRVAHMTGLKRLDNLHRLAVVEDGVGDIETGVWRGACITARKGLN  
ATGQADRTVWADSFQGFPELTGSDHPLVEIDLHOYNEAVDLPSTSETVRENFARYG  
LDDNVRFLAGWKFDTMPAPVAKQALVMRLDGDYGTATMDVLSILYERLSPGVYIVD  
DVCIIPACREAVHDFRDLGIRPTIHRIDQAGVWRHSG"  
4897..5514  
/gene="ty1J"  
/note="orf19\*"  
4897..5514  
/gene="ty1J"  
/function="synthesis of NDP-deoxyallase during tylosin  
production"  
/codon\_start=1  
/transl\_table=11  
/product="putative NDP-hexose 3-epimerase Ty1J"  
/protein\_id="A041820.1"  
/db\_xref="GI:5305789"  
/translation="MRDTRRPLGIEGAVYIOPEIHPDRGEFFHAMFOSSEPRRTGHS  
FSPYOVNIAVSRKGLRGTHFSVPVPGAKYSACVAGAVEVYVDRLGSPYGGWRA  
VPDEYNRTAVVYPACLGAFVALTDRTTLVYLCSESYARREHAYNPDLPLATAMP  
EDVELLSRDRAPTLAEARKGILPSTQETREHGLPEHARR"

BASE COUNT 799 a 2072 c 2098 g 864 t

ORIGIN

Query Match 7.9%; Score 110.6; DB 2; Length 5833;  
Best Local Similarity 47.4%; Prid. No. 1,8e-13;  
Matches 397; Conservative 0; Mismatches 434; Indels 6; Gaps 2;



OY	466	tgagatcccccagcagccgacgttctgtgcgttgcgccaacaaagtgtgtgcattgcgg	525
Db	3490	tggacaccccccacacacagagcgtctccggccacattcatcccccagattccggccgc	3431
OY	526	tgtgtgtaagctctggagaacccgattccaaagatctggtcctgtcgtgtgtatctgaagcctgc	585
Db	3430	gggtgtccggagctgtgcgcctctcccttgaaacacatcgtcacccggcctgtgatatcactca	3371
OY	586	gcccgcgaag---acagtcaacttccacagagacttgcgcgaacccctcccaatacga	642
Db	3370	ccggcccgccggcgaacgaagggagatctctgtggccgacttgcgcgtctcccaatggcgaacccag	3311
OY	643	tcttcattgtctgtctcgcaagtgltacccgaaagatalccgcgaacttgaatactaaag	702
Db	3310	tgatcttcccccgtctgtcgcacatccctctacgagaccggagactcttccagagcgacac	3251
OY	703	atacgaatgacccgttccggaatgagatataccttcgaaagtcgcaagaagcgtctacg	762
Db	3250	amcagccaccccccgcgcgcgcgcgcgcgaagagcccttgagggcctgttgagctgtgcgc	3191
OY	763	actatctgataccgatactcatctagcaaacgcaagcagaagccggaaacgcagacttca	822
Db	3190	actactctgcagccggctgatcagcggcgaacacccggccggaaatccgcgcagcggatgtctgc	3131
OY	823	tcgttgcgaacgcccgaagtgtcaatgtggtcgacacataccagtgaagaagccaagagatgt	882
Db	3130	gcacacatgtgtggcgacagcccgcgcgtgtgcggcgtctccgcacccacactctctggacac	3071
OY	883	gtgagcctgttactgtgtgcgcgtctgatacagtggttcaatttccatgccttcaagata	942
Db	3070	cggagctctcgtgctggtccgcgcgcgcgcgaacacacacggccacgatagtatcagatgagcgtgc	3011
OY	943	agttcctgtgcgaagaagcccggagatcgtgcacagagatctatcagacgttcccgagcgtatc	1002
Db	3010	tcgattctgcacacacaccccacggcctgtgcgcgaactacacgtgatctccggctgtgtgc	2951
OY	1003	cagcgccttgcgaggaactactcctcggcgtcttgcgtgttgccagt--ggccgatacc	1059
Db	2950	cgggcgcgcgtgtgacgaacatctctgccttatctgtccatcccccacggcgtccgcgccttcgc	2881
OY	1060	tcaacttcgaattacgaagtttcatatggtgtgcaacttgaagaagaagtgcacagatactctgtac	1119
Db	2880	ccacccggcgacatcgaatgacgaagcgaacacacatccacatccggcgcgcgcctgtgattcc	2831
OY	1120	cgcgaatgctgtctgtgcctcgtgtatgtagcgtgaagaacgcttgcctccgaatgcgtacttca	1179
Db	2830	tgtgtgtccggcgcacacccggagagaagcgtcttctccgaacccgagagcctttccacatcc	2711
OY	1180	gtgcgcaaaagtttcaacaacacacttgcgcgaagcgaacatctgtgtcttgcgcgc	1239
Db	2770	acggcgtccggagagcgcgaatgtgccttgcgttgcgttgcaggaacccacacagtccttgagacaga	2711
OY	1240	acctgtgcgcgcgggaatcatctgtcaaccttaagaagtgtgtgaccagaattcctgt	1296
Db	2710	acctgtcccgatgacatcgaatgtgccttgcgcgcgcgcgtgtgtgagagcgttcgcgc	2654

RESULT	9
AF055922/c	
LOCUS	
DEFINITION	AF055922 12905 bp DNA BCT 04-MAY-1999
	Streptomyces fradiae dipeptidyl carboxypeptidase (ddca), tylosin
	resistance protein (tlrB), glycosyltransferase (tylN),
	methyltransferase (tylE), 4-ketoreductase (tylD), ferredoxin
	(tylH2), cytochrome P450 (tylH1), microcin O-methyltransferase
	(tylF), epimerase (tylJ), acyl-CoA oxyase (tylP), and
	butyrolactone receptor (tylQ) genes, complete cds.
ACCESSION	AF055922
VERSION	
KEYWORDS	
SOURCE	AF055922.1 GI:4210925
ORGANISM	
	Streptomyces fradiae.
	Streptomyces fradiae
	Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:

REFERENCE	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
AUTHORS	1 (bases 1 to 12905)
TITLE	Fouces,R., Mellado,E., Diez,B. and Barredo,J.L. The tylosin biosynthetic cluster from Streptomyces fradiae: genetic organization of the left region Microbiology 145 (Pt 4), 855-868 (1999)
JOURNAL	99235571
MEDLINE	2 (bases 1 to 12905)
REFERENCE	Fouces,R., Mellado,E., Diez,B. and Barredo,J.L. Direct Submission Submitted (27-MAR-1998) Ingenieria Genetica, Antibioticos S.A., Avda. Antibioticos 59-61, Leon 24080, Spain
JOURNAL	Location/Qualifiers
FEATURES	1..12905
source	/organism="Streptomyces fradiae"
gene	/db_xref="taxon:1906"
	complement(235..1431)
	/gene="ddca"
	complement(235..1431)
	/gene="ddca"
	/note="Similar to Bacillus cereus alkaline D-peptidase and to Streptomyces serine dipetidase"
	/codon_start=1
	/transl_table=1
	/product="dipeptidyl carboxypeptidase"
	/protein_id="AAD12161.1"
	/db_xref="GI:4210926"
	/translation="MRKKVYATLMAVTAVIGVTGQAAPRTAGCTAECAGHRCPTALL OREADLLRACPGVLADRTPRTVTVVRSGFDTDGTPEPKARRISLTFTPA ATVALQVGESGLSDDTVDHMLPGLVSGRKDGSAITVGMILRLTSGLRVLPPLPYI FLEHYGNRRRTVPDQAVRALRHPPVAHPGTRMGYSNTNYILAOMIVEKAIGHSM QHREYRVIEPLRGHRHTVADPTPEPIGPNAHVGYREFEKGLEADPADPMGEVDVT VGSNMGAGAEMISTETEDGNRFLLALGGVLRPAFLAEKKTKVRAGEFDSAMPVR YGELMAPNSCGGVWFHGDIPEFKIRNGVTDGSRNSVNVSTQSMTVEPCVPARP GRRRP"
gene	1788..2630
	/gene="tlrb"
	1788..2630
	/gene="tlrb"
	/note="Similar to Micromonospora griseorubida Myra protein"
	/codon_start=1
	/transl_table=1
	/product="tylosin resistance protein"
	/protein_id="AAD12162.1"
	/db_xref="GI:4210927"
	/translation="MKKNVRYLCRPCAPLRASDDTLRCENGHTDVARGCVNLL RRPRKLADTTDMAARAALDSGHVAFLETRHLTGRAAGAACAPCDVDIGGSTGH HLAAVEEFDEAGILLDMDSKPAYRARARNPRASSVVADWADLPLRDGAAMALNV FAPNRPEIRRIILRPGTLLVTPQQOHLAEIVDALVLRDRLKEGLAEOLAHDFEE AVOGCEIRLTRLDHALGRVYVAMPSPSHQDPDELARIALRELPGIHETVLSVFVC RPDP"
gene	complement(2700..3968)
	/gene="tyn"
	complement(2700..3968)
	/gene="tyn"
	/note="rhannosyl transferase"
	/codon_start=1
	/transl_table=1
	/product="glycosyltransferase"
	/protein_id="AAD12163.1"
	/db_xref="GI:4210928"
	/translation="MRTAILTMGRGDVOPFVALGTCRLARGHEVYLGADEALPLVE OAGLEYRATRODPGFEFTTMEPVETTLLRGPMARDLMALPPAPEYDOEVLDRIERAG EGVDLVHADLVTTTALGESPTWLSYNMMPNSTWTFPVESSQRMGLTPLYNL THMAEREEDMGWRANERGRGRGLRPFKSSPLRLGHRPLLYPFSVSLKRPD WPCCGCHTYGFMDQPMRPSDELEDLAGERPVLTLTGSTMPVNHQOEAMEYAAVA ARGRRRLLYVGPEGALPDGLRVPDADYSMLMPRTAAVHHNGEGESTADAVRAGVP OVLYRVADHDFAARLRIMGTARPRPLARMREALAASYRTAAYTDPAAVARRLG EAVAAREGENACTYLIEMAETRTTAHTPC
gene	complement(4014..6229)
	/gene="tyle"
CDS	complement(4014..5201)

```
/gene="ty1e"
/notes="similar to Streptomyces mycarofaciens MdmC protein"
/codon_start=1
/transl_table=11
/product="methyltransferase"
/protein_id="AADI2164.1"
/db_xref="GI:4210929"
/translation="MAVQKEATLVROIIIRAAGGHADVRELVAEHGPEAVTAVLDEI
VSRAPEPNDVPLVLAVERSGDALVPRRLAVALOAGAPVRAAPDDGCFVAMRYEELD
ELVRELFGPCREBAAGTGGTTLFPYATSGMHIDYELAAOATATVLACGTSAPKDL
NELTSRYLTPKPKSLHMPHYDRHREVRNEVRVLEIGIGYQHPENGGCSIRMKK
HFFHRLITIGLDIEDKSHAEBQRTITVGDNDPGCLTELARIGPFDIVIDGSHIN
EHVRSFHALPFPVPRGGLYVLEDLMTAVMPGSGSDPSKSDLTSLGLVKSIVLSQ
HQLPDEDSRSPGYADRHVYGHVHNLAFIEKGVNSEGIPGWIIPRDPDALVAASSG
GAA"
complement(5219. .6229)
/gene="ty1D"
complement(5219. .6229)
/gene="ty1D"
/codon_start=1
/transl_table=11
/product="4-ketoreductase"
/protein_id="AADI2165.1"
/db_xref="GI:4210930"
/translation="MRAAGDWAGRTVVTAGTALGIGSHYAERIALGALGARVYGLHRSAR
PAGAPAPAPAGRSCLVRADLCDEAATRAFCENAPGADVIVHCAGLGNQYKRDI
SASVLDANRGTAHVLTARDGAGAVLLSSTEYVCAPRSPAREDEELRRVVPAG
NTEVLSKICEIAELHGAEPGSRIFRVRGNYPYMGNGGRTVITISMAVRAAG
ETLEIMGRCORRSRTFHVHVDIVNATLRVNEADKPYPMNGGTPREVSILDLARVYSAV
GGKGIKRTYPDKPSPPPAQRDLISRMKDVTCGEPRLRVGIETQITIMYRGHHRPSAP
HS"
complement(6226. .6471)
/gene="ty1H2"
complement(6226. .6471)
/gene="ty1H2"
/codon_start=1
/transl_table=11
/product="ferredoxin"
/protein_id="AADI2166.1"
/db_xref="GI:4210931"
/translation="MVRIDTGRGCGAGCGCCERAPVRFRODEGVGVLDKRTPPRAVW
BEVREABDLCPRANVLLSGDTGAGAGAAPPTGRDA"
complement(6507. .7769)
/gene="ty1H1"
complement(6507. .7769)
/gene="ty1H1"
/notes="similar to Streptomyces griseus Soyc"
/codon_start=1
/transl_table=11
/product="cytochrome P450"
/protein_id="AADI2167.1"
/db_xref="GI:4210932"
/translation="MSSGSDARPSOKGILLPAARANDPEAGRSIAMPVARTCPES
PBEQYALAAEERPIARAEIMDGPVYLISROHVRLLADPRYSIHAPLRLSPSDG
EBAESKSLTLPDPHGLRGHIFPEFGKRVKDVRSVEQVLTGLDLDTAKGDAD
LADFAALPMAQTQVLCRLDIPYEDRDYEQERTQAPRAAGEALTEALLDELVDRL
ISGKTGREGSDMLGSMVAOARGGGLSHADVDNALLAAAHETTAASVPTASVIVLL
OHPTAARLELVNPGLLPGAVDELRLYLSADGIRRAATDIEDGTITAGGGLVALL
AAANRDEAVFSEPEAFDHIRSARHVAFCYGHQCLGQNLAMLEVALGAVLERLPA
LRPTDVAGLRKLSASAVTCVYELPVAW"
8033. .8800
/gene="ty1F"
8033. .8800
/gene="ty1F"
/notes="similar to Micromonospora griseorubida MycF"
/codon_start=1
/transl_table=11
/product="macrocin O-methyltransferase"
/protein_id="AADI2168.1"
/db_xref="GI:4210933"
/translation="MAPSPDHARDLYIELLKRVSNVYIEDPPTHVAGMTTDSFQRTS
RSGEDPTVAHTMIGLKRLDNLHRLADVDEGVGDSETEGVWAPCIETARGLLNA
YQADRTVWVADSFGFPELTGSDHPLDVEIDLHVEAIVDLPTSEITRENFARVGL
```

```
LDDNVFLAGNFKDTPAPAPVROLAVMRIDSDYGTATMDVLSYERLSPGGYIVVD
8837. .9454
/gene="ty1J"
8837. .9454
/gene="ty1J"
/notes="similar to Streptomyces glaucescens Strm"
/codon_start=1
/transl_table=11
/product="epimerase"
/protein_id="AADI2169.1"
/db_xref="GI:4210934"
/translation="MNDTTPRIGICGAVVIOPEIHPRKRGEEHANFOSSEFRRLTGS
FSVPVYNIAVSRKGAAGIHFSESEVPGQAKYACVQAGVEVYVDIRLCSPTYGOMRA
VPDEVNRAVVPAGIGRAFPALDRITLVYLCSEYARREHNVNPLDPLATAMP
EDVELLSRDRTPAPLTAARRGILPSTQEHREHNGLPREHARR"
complement(9615. .11591)
/gene="ty1P"
complement(9615. .11591)
/gene="ty1P"
/notes="similar to Rattus norvegicus pristanoyl-CoA
oxidase"
/codon_start=1
/transl_table=11
/product="acyl-CoA oxidase"
/protein_id="AADI2170.1"
/db_xref="GI:4210935"
/translation="MPLVQAPRRPLCYGGGSHVIAHAPQAPGHITPDHSSPDADPIA
GSAFPTASITDQDTPASRLTNLIEDRDRTRLKGPWNLITTEFRHRRBGLPAARLT
ALAYDLRFVNELVDEPABELARDVRLALSHMTGVDAQOTLAGIHNLNLGLSLAD
HRSDEKDISEFLSMRTGTFELCTELAHGNDAAALRTAEIDRDKFVNLNPTPGAQ
KEMPNTSTGGPRTALVAVRLVVDGHDGVFLPLSPAGTLGIRLRRLPERTGH
PYDCHTSPDHVLRPEALLBGVHGRLAGSGESFPLRHRHRRRLLEGLADAAMPTLS
AAVCVTPRALSLAVRYSHRHISASRQRCPTPLARHRSRRRLLEGLADAAMPTLS
RRALSRMAGRDETDQADARVLAIAKMTTWQARAVTIECRRCAGQCLFPVNGLADH
PLNLEGTITABEDNLTWLKASSEMIFHRTQORPLDGPPEQSLTDPRFLHDLAD
ABGLWREKARAGVROGSPDGLRWNGASAAVEVRAVAVQAAGSILEAAGHADGA
APROLVWLSRLFMLRHLAPHTGDLAEGRMADVRLALPGSVDEVIAGLAPHMNTLV
EAPDLPGEFGLGPILANAHETRPQIVDTLLSA"
11938. .12612
/gene="ty1Q"
/gene="ty1Q"

Query Match      7.7%: Score 107.4: DB 2: Length 12905;
Best Local Similarity 47.2%: Pred. No. 7.9e-13;
Matches 395; Conservative 0; Mismatches 436; Indels 6; Gaps 2;
```

QY	466	tggtatcgccgcgagcagcgccaggttcgtygcctgycgcaacaagatggttgcatgcggy	525
DB	7432	TGGAACCCGCCCGGACACGAGCGCTCCGGCCACTTCAATCCCGAGTTCCGGCTCGGC	7373
QY	526	tggtgataagctggaagaaccggaatccagagctggtcctgctcgatcgagagctgc	585
DB	7372	GGGTGCGGGGACGTGCGCCCTCCGTGACACAGATCGTACACCGGCTGCTGGATGACTCA	7313
QY	586	gcccgaag--acagtgcaactcaaccgagactcagcgaaccctccgatacgca	642
DB	7312	CCGCCCGCGCGAGCAGAGCGGATCTGTGGCACTTGGGCTCCCATGCGGACCCAGG	7253
QY	643	tcttcattgctgctcgcaaggtctacccgaagaagatatcccgcaacttgaatatcacta	702
DB	7252	TGATCTGCCGCTGCTCGACATCCCTTACAGAGACGGGACCTTCTCAGAGACGGACCG	7193
QY	703	atcagatgaccgctcgagatgagcagataccttcgcaagagcgcaagggagcgctacga	762
DB	7192	AACAGGCCACCCCGCCCGCGCGGCGGAGAGGCGCTGGAGGCGCTGCTGGAGACTCGCG	7133
QY	763	actatctgataccgatacagcaacgcaagcagaagccgggaacccgatacga	822
DB	7132	ACTACGTCGACCGGCTGATCAGCGGCAAGACCGCGGGAATCCGGCAGCGGATCTCTCG	7073
QY	823	tcgttgccaagcgcaagtcacatcggcagcagataccacagtcagcaagccaagaagatc	882
DB	7072	GCAGCATGTGTGCGCAGCGCCGCGGTGGCGGCTGTCCACCGCCGACGTCCTGACACAAG	7013

[illegible]

RESULT	10
AFI27374	
LOCUS	AFI27374
DEFINITION	AP127374 63734 bp DNA BCT 27-MAY-1999 Streptomyces lavendulae Lina homolog, cytochrome P450 hydroxylase ORF4, cytochrome P450 hydroxylase OFB3, MltR (mlt), Mts (mts), Mtr (mtr), Mto (mto), Mtp (mtp), Mit (mit), Mtn (mtn), Mtm (mtm), Mtl (mtl), Mtk (mtk), Mtu (mtu), Mti (mti), Mth (mth), MicG (micg), Mtr (mtr), Mite (mite), Mld (mld), Mtc (mtc), Mtb (mtb), Mta (mta), MmcA (mmaa), Mmb (mmcb), MmcC (mmcC), MmcD (mmcD), MncE (nmce), Mmf (mmf), Mmg (mmg), Mmh (mmh), MmCI (mmci), MmcJ (mmcj), MmcK (mmck), Mmcl (mmcl), MmcM (mmcm), MmcN (mmcn), MmcO (mmco), Mrd (mrdd), MmcP (mmcp), MmcQ (mmcq), MmcR (mmcr), MmcS (mmcs), Mmct (mmct), MmcU (mmcu), MmcV (mmcv), Mct (mct), MmcX (mmcx), and MmcY (mmcy) genes, complete cds; and unknown genes.
ACCESSION	AF127374
VERSION	AF127374.1 GI:4731328
KEYWORDS	
SOURCE	
ORGANISM	Streptomyces lavendulae. Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE	1 (bases 1 to 63734) Mao,Y., Varoglu,M. and Sherman,D.H. Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from Streptomyces lavendulae NRRL 2566 Javandel Chem. Biol. 6 (4), 251-263 (1999)
AUTHORS	JOURNAL MEDLINE 99201491
REFERENCE	2 (bases 1 to 63734) Mao,Y.O., Varoglu,M. and Sherman,D.H. Direct Submission Submitted (10-FEB-1999) Microbiology, Uni. of Minnesota, Box196, 420 Delaware St. SE., Minneapolis, MN 55455, USA 3 (bases 1 to 63734) Mao,Y.O., Varoglu,M. and Sherman,D.H. Direct Submission Submitted (27-MAY-1999) Microbiology, Uni. of Minnesota, Box196, 420 Delaware St. SE., Minneapolis, MN 55455, USA
TITLE	
JOURNAL	
AUTHORS	
REMARK	Amino acid sequence updated by submitter
FEATURES	Location/Qualifiers 1..63734
SOURCE	/organism="Streptomyces lavendulae"

CDS

/strain="NRRL\_2564"

/db\_xref="caxon:1914"

123. 1367

/function="dehydrogenase"

/note="ORF6"

/codon\_start=1

/transl\_table=1

/product="unknown"

/protein\_id="AAD28446.1"

/db\_xref="GI:4731329"

/translation="MRNTLFELGDIPIPIGTVPEQMAVYIRREYQGHQPHAFSEVVD  
VPKRGQALVLVMAAGINYNKNWASLQGPVDYISAKQKGSHEDEHIGSGESGVVA  
AVGSEVTVAAAGDEVILISGQCMTEFLADIRLGADPMYSSQSWGEGNGSPAGQAL  
VDDYQKPRKPGLTWEEACPLLTGATAYRQLCGMPHDVYRDPVYLWGAGAGLGS  
AIQITRARGGIPVAVVSEERARCYCRELGAQGTINRDPFHMRPLDIGHEMGMRT  
EGVRAFGRRFMEVLEGRSPRIVLHESQATIPISWLCNAGAVYIGGTYGYNDI  
DRELFMQRKQSGSHFANRLRCQCDYIHMVANQGLDPCLSMTGGFDDIGKANQLMND  
OHQGNQAVLVNAPRTGLTPPA"

1428. 1958

/note="ORF5"

/codon\_start=1

/transl\_table=1

/product="Lina homology"

/protein\_id="AAD28447.1"

/db\_xref="GI:4731330"

/translation="MDKLIDILMSEREIRAVLQRYCGLRDELYKSAAYHEDAHDR  
GVIRNNAIDPVKQIVPLRLDAYVTGTLTGHSTETIDGDAVSEYCTAAYHRESGI  
KRVEQFACRYVDFEFRDGVKIKARLVLDNDFSLAQCEVPLDPANAQAGFNSHRDLTD  
ASYOVLPRGPDAPTL"

2006. 3229

/codon\_start=1

/transl\_table=1

/product="cytochrome P450 hydroxylase ORF4"

/protein\_id="AAD28448.1"

/db\_xref="GI:4731331"

/translation="WSDTEQHAPTLPRQRTCPSPPELLELRKTDPISMRRADSP  
GULTLRHADVRAALADPGVSSHKAPQPMRNILAPEMRASHYLPGLIFMDPDHRY  
RLTLTKRMTMAIRKLEPRLEQIVETLDAEAGQTVLDVQSFALPIPLVLICELHG  
IRYERSEEFMDVLRALMDATPEELALCARMEFYMKLAAKRAAPGDDLSELIA  
DDPADAPLTDLIEIGIGVLMILAGHETSNANLIGVGTITLENDDOMALINDDISLDR  
AVELLRHQITVQOGLPGCVTRDMEINGAHQKTESLALSLPANRPAVFPDPDLD  
ITREINPLRAGHGHIHLCGLKELARVENRQAMKQLVTRFGLRMAAAPEDIRMRDQI  
YGVGNVLVPTWDEAK"

3226. 4458

/codon\_start=1

/transl\_table=1

/product="cytochrome P450 hydroxylase ORF3"

/protein\_id="AAD28449.1"

/db\_xref="GI:4731332"

/translation="MTCGEAAVRCPCGACGAPAYPEHADRLLEDPYWEPLRRERPL  
ORVTLPYGCEAMLATRYODVAVFADRRFSROLAVAPGARPLPHOPPPAVLSVGP  
DHATRLRLVGVTFPRPRYEDMRPLIORTAGSLDAMEMKRPADIVDEDSLPAVSMI  
CELLAVPPEDKRRFCVMSDALVLLTTHAIPKAVORYKQMDYVLAGVLQAKRVAPVND  
IGSLVTADEEDKLTGELGLALAEALILAGETASQIIPAEVLFRHQLLERIND  
HDLPLDAVEELRLRPVIGTVGPPRTATEVLELGGVLYVRAGEVYVSMGAAANDPELF  
TDPDLDLARRPNPLHFGGAPRHLCQAQLARVELQITLTTLPRRYPRILAVBESLS  
SMKEELMRGHTMPVTM"

4552. 5658

/function="esterase"

/note="ORF2"

/codon\_start=1

/transl\_table=1

/product="unknown"

/protein\_id="AAD28450.1"

/db\_xref="GI:4731333"

/translation="MSTIDWEHSTKESAGDPAALRRLLTDVVARGCAQALCYMRGCT  
VLWDSFECSSDSLFLVLAATKPAALALVALAEKGLIGDRPAVEYVQFAGAGSD  
VTVRVILDHARGAVPGVGRIVRTKATGADWESVRLDOSPKRKGDEVAAVHMSGCD  
ITGELVQVTVRSPRDRVYTSLEFAPGLINDLHGLPESANPRVYPARAHPSPMPQW  
MSNRGIGQAYITPSNGLSGTRAQMARITQMLMBGSLDGILRVLPETVEARKPSNDG  
GIDSLKRPVMSHGFMLGGGPGDPRLSVNLGTSDFSPSGFHAGTTSSVYADPTRE  
LVLVLSNIQPEFAGIGIERLEVSALACGCEAG"

5992. 6849

CDS

**CDS**

```

/strain="NRRL 2564"
/db_xref="taxon:1914"
123. .1367
/function="dehydrogenase"
/note="ORF6"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28446.1"
/db_xref="GI:4731339"
/translation="MNTLTPELGDIPDICTVPEBQVYSKAKQSGHSDPHICGSSGVV
VKVSGCALVLYMAAGTIVNYMVASLQGPVDYISAKQVQSGEDPHICGSSGVV
AVGEQTVQVAGDEVLTISCCQMTETADIRLGADPMTSSQSQSVGEGVGSAPFAL
VDYQCHNPRL/TEWCAEFLTGATAA ROLCGMPDHVRPDPVLINGAGGIGLSM
AIOIIRABGICIVAVYVSDDEERARVCELGAAQGTINLDPDHMRGLPIDHMGKMT
EVSARAGRRFMEVLEGRSPRIYVLEISGQATPTSYLCDNMGMYICGGTGYANDI
DLRFLMRKQRLKSGSHSRFALROCRDVIHVAHQDLPCLSMTCGFPDICKAQLMDND
QHPQGNQAVLYMAPRIGLTTPA"
1428. .1958
/note="ORF5"
/codon_start=1
/transl_table=11
/product="Lina homolog"
/protein_id="AAD28447.1"
/db_xref="GI:4731330"
/translation="MOKLIDLWSEEREIRAVLYQRYCHGLRDLDELYKSAVHEQADSP
GVIRGAHNPVQVQIYVLDLADATGTTLHTLHGSTIEIDGAAQVESCTAVHESDGI
KVEQDAGVYVVEEERDQGVKIRLVLTNDSLAGEVLPDPAEQAGFNPSHRLDLD
ASYQVPLNGPDPAPLT"
2006. .3229
/codon_start=1
/transl_table=11
/product="cytochrome P450 hydroxylase ORF4"
/protein_id="AAD28448.1"
/db_xref="GI:4731331"
/translation="MSDTQHAFTLPKQRTCPSPPELLELRDPDISMRADSP
GNULTIRHADVRAALADPGVSSHPKAPQPMWRNLAPMRAREHYLPGLIFMDPDDHRY
RLRLTKRPMKRIKLEPRL EGVETLDAMEAGGTVDLYVSFALPIPLPLVCEIAG
IYEEKEEPMDFMRLQALADAPDEELGALGARNEMPMKLAAR RANPGDGLSLHAG
DDPADALDLDLACGIVMTLADGETSANMLGCVGYTTLLENADQALRLDQISLDR
AVELLRHQTYDQGLPCRTVDMETIAGHGVMTGSESLASPRANDAPVFPDPDLD
ITREHNPILAFSGHILCLGEMELAREHMQAGLVTRPGLMAALAPEDIRWDQI
VGVGVNLPVTEWDEAK"
3226. .4458
/codon_start=1
/transl_table=11
/product="cytochrome P450 hydroxylase ORF3"
/protein_id="AAD28449.1"
/db_xref="GI:4731332"
/translation="MGPSPAAVGCSPFGAGEAPAYFPHADRLPDPVMEPLRERPL
GVNLTAYGGEAMLTAQYVOVRAVADRRFSROLAVNPGA PRLPHQPPDPAVLSVSG
DHARLRPLVYKQTPRPRVDDMRPLQRTADGLLDAMEBGRPADLVEDSSLPRAVSMI
CELLVAPRPDKRATFCVSSDALITTAHTAQAQRDQIMQMDHIDVGGVLAQRRARVPL
IGSLVAPROBEKLTGEGELVLAELAELLVAGYASDQIAPNFVLVLRPHQLEIRIND
HDLIDPAVELLRLPVIIGVDFPRATDDELVEGLVAGGEVYVSSKAAANDPFLF
DPDELDLARRNPRIKFGFAGPHNCLGAQ LARVELITLTTLPRRVRLRAVBEESL
SKKEGLMVGSMTPMTVM"
4452. .5658
/function="esterase"
/note="ORF2"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28450.1"
/db_xref="GI:4731333"
/translation="MSTIDDEHSHSTKEACMDPAALRLRLDVVYRAGCAQOLCYMRKGT
VIVDSFGGSSGSLPLFVYATPRVALAYNALEERLILDPRLVAEYQFAPRAGSGT
VIVRHLVAGSACVAPGRCIVPMRTAGDMERSVRDLEQSRPKMVGGEVVAHYHMSGF
ITGELVRYTQSGSFSDPYVSELFARGLNDLIMGRLGSGMWPRHVRARAHNPSEMPOW
MNSNKORTQAVYLPSCGLSTQAQMAQRFQMLDEGGLDGIIVLRPTEVEARPNKDE
GIDASILKRVAVRSHKGLGGPGDPGRGLSNVIGRTSYDPSAFGHAGTTSVWADPTRE
VLAIVLSNQDPFGAGIERTLRREVSLDALGCAEAG"
5992. .6849

```

5992.6849

/function="transcriptional activator"  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AAD28451.1"  
 /db\_xref="GI:4731334"  
 /translation="MLNLPKGMERAHPHSPPOVGLIPLEVRSAAGAGCTGANVSGIV  
 RLKLLAATLRQAQAMSTERILKEVWADMPATIDKRAVAVALRLRVLGDNQGRMLLT  
 RSGCVLDLPDHLDAVRAETLVREGRAALACDPYVAARILTRALDQMGREPADANM  
 ASIVSQRTELENRSEAVOAHIDARLELGHOBELGELSRVPLGMRSEBDLPFPDPA  
 LALYSGKQAEALAYMQDLQALAMENLIGIDGRQLDELHLRLILADAGMLTGSFAP  
 AEPILVRS"  
 complement(6854. .7666)  
 /gene="mitr"  
 complement(6854. .7666)  
 /gene="mitr"  
 /function="aminoquinolate dehydrogenase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="MitT"  
 /protein\_id="AAD28452.1"  
 /db\_xref="GI:4731335"  
 /translation="MEFLPSPAAGVSGATRLYAVLGDPAQYKAPGLINPLLSBSGLDA  
 VYVPHVARRDLAEYVEGILKRIGNLDGLLTVPHKALCGIADLGPRAALIGTANNM  
 REEPGRKWAENFEDLGFVOQJQAAGHTVNRHVALVGAAGSASITATLMDAAVSV  
 SHVDIWRQALDLRLTIGSRRPDGLRALGPDLLAADPAFVANATPLGMRSEBDLPFPDPA  
 RRPAPVAVDVYMKPHEFTALLSAATAGRAVHHGIIHMLEQVPCYRAAFGWP"  
 7895. .8842  
 /gene="mits"  
 7895. .8842  
 /gene="mits"  
 /function="kinase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Mits"  
 /protein\_id="AAD28453.1"  
 /db\_xref="GI:4731336"  
 /translation="MTGDTDGAGGGDYTFRMPAAGDVTADLLLAARYGILLGHREDP  
 LAGVAGWAMICDAAGVTWTPRGPSMAGMLITAAFGOLLPGTFVACADDDLAIALE  
 SAAACRHLILYGVCTGJGGGIVHGRAMPGRGSCGVHVVVDGSRGDCGRACC  
 VOAAVSGPTELRRAERRGRETGEDELASGARLIAPMAEAVIDESAAALATAVNGIOE  
 LAHPELVGVGSGFAGVGYVAASVAHYERTLTRETDPYVRPAVILGSRSLHALLL  
 AEAHGRGRPEPSRVSSDVSVFSGGVTDRAVGRSD"  
 8916. .10460  
 /gene="mitr"  
 8916. .10460  
 /gene="mitr"  
 /note="similar to mitomycin C oxidase MctA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="MitR"  
 /protein\_id="AAD28454.1"  
 /db\_xref="GI:4731337"  
 /translation="MLDRSVYTRVGASAAAANAAGRATGATAAVGLPQVGRASGAS  
 VWMASIRLSELVLPADGTGERAKRLYSQFDFGIRQAVAYCRTEEDVKTTLAEQO  
 DHALPTIPSSGSPFGYSTDGIVLDSGFHASLIRNTVVMGAGTQOVDALPALS  
 RGVAAVSGACVSGCFYVOGGGLGMOGRKFGMACDLVSARVYLADGRAVTAATAS  
 PLFPMAMRCCGGNGFVYTFELRPTDVSIVSNLTPMPSARVILEAMOHVILIDP  
 RLGAAMRCCGADCTGTPVYVYVYAGMIGADALTPVDSVAVSGSARATRSKAKLS  
 OHDDAMAYGCADLTPEOCHTVYGSPEALFRPNFSMDRNLFSRAIGQGVETIKLS  
 FAADPAAGQFRLLSFALGGAANRPDRTTTAYHRDVEFYLGFSIGLNDPEYTLAEDDR  
 LGRANNAARLTLDPHSNGESYONFIDPELDWKSAYVAENYVRLAAVKAAYDPHRL  
 SFAOV"  
 10607. .11101  
 /gene="mitQ"  
 10607. .11101  
 /gene="mitQ"  
 /function="putative regulator"  
 /codon\_start=1  
 /transl\_table=11  
 /product="MitQ"  
 /protein\_id="AAD28455.1"

Query Match	7.4%	Score 104.2;	DB 2;	Length 63734;
Best Local Similarity	47.4%;	Pred. No. 3.1e-12;		
Matches 452;	Conservative 0;	Mismatches 483;	Indels 18;	Gaps 4;
QY 452	cattccacactcgatgatgatcgccgcgcgagacgagccagtttgcgtgtgccaaccaagt 511			
DB 2272	CTTCCGTGATCTTCAATGAGACCCGCCGACACACCCGCTACCCGCCCTGCTCACCAATG 2331			
QY 512	ggtttgacgtccggtggtgtgtaagtcttgagaaacggagatccagagagcgctgtgtct 571			
DB 2332	GTTCACCAATGCGGGCCATCCGCAAGCTGACACCCAGGATCGAGATGTCATACCGAAGC 2391			
QY 572	gatcgagagcctcgccgcgcaagagcaagtgtcaacttccacgcagaga---ctacgcgaacc 628			
DB 2392	CTTCGACGCGCAATGAGAGGCCACGAGGGGGCACCGTCGACCTGGTGCAAGTCCTTCGCCCTGCC 2451			
QY 629	cttcccgatacgatcttcaatgtgtgtctgcgcaagtttccagagaagaatatlccgcactt 688			
DB 2452	GATCCCGGCTGCTGATCATCTGCGACCTGATGGGATCCGCTACGAGGAGCGCGAGAGATT 2511			
QY 689	gaataacctaaagatcagaatgaccgcgttcgcgagtga---gcatgaccttcgagagagc 745			
DB 2512	CATGNCATGATGCTCTGCGACCTCCAGGCCCTGAGACCCACAGCCGAGAACTGCGGGCCT 2571			
QY 746	caaggaagcgtcctcagactatctgataccgatactcagacaacgacagacgaagacgag 805			
DB 2572	CGCGCGCAGAGATGAAAGAGTTGATGTAAGCTCGCCCGCGCAAGCGCGCAACCCGCG 2631			
QY 806	aacgcagctatcagatcgtttgccaacgagccaggttcaatggcgagccgataccagttga 865			
DB 2632	CGACGACCTGCTCAACGACACCTCGCCACAGACCCGACCGCGGCTCTACGAGACT 2691			
QY 866	cgaaagccaagagatgtgtgtgtcttacttgttggcgagccgtgagataagtgatgaatt 925			
DB 2692	GGAGATCGCCGCGCATGCGGCTGCTAATGCTCATTCGCGGGGACGAGACCTGGCCCAAT 2751			
QY 926	cctcagcttcaagatgagttccttgcgcaaaagcccgagacatcgcagagagctatcga 985			
DB 2752	GCTGGGCGCTCGGACACTACACCTCTGGAGAGACCCGACGAGTGCGGCCCTGCTCCGGA 2811			
QY 986	gcgtcccgagcgtatltccagcgcgttgcgagaactactccgcgacttctgcgtgtgtgc 1045			
DB 2812	CGACATCAACCTGATGAGACCGGCGGTGAGAGAGCTGTCGCCACCAAGACATGTCGA 2871			
QY 1046	cgatgagc-----cgatcttccactccgattacgaagtltcatgtgcgycgaactfgaaga 1099			
DB 2872	GCAGGGCTGCGCGCGCGGGGTACCCGGGACATYGGAGATCGCCGGGACCAAGGTAAAGC 2931			
QY 1100	aggctgacagatcctcgtctacccagatgtcgtctgtgcgttgaatgagcgcgaanaagcgtg 1159			
DB 2992	CCCCGACCGCTTCGACATACAGCGGAGAGACAAACCGGACCTCGCTTCGGGACAGGAT 3051			
QY 1220	ccatctgtgcttgcgcagcactctgtgcccgcgcggaataatcatgcatccctcaagaagt 1279			
DB 3052	CCACTCTGCTCGGCGCATGAGACTCGCCGGGTGGAGATGCGCCAGGGGTGGCGGGGCT 3111			









Accession	Version	Keywords	Source	Organism	Reference Authors	Journal Medline	Features	CDS
QY 935	cagcaltgagagttctctgcccacaaagcccggaagcaltgcgcagagaagctgatcagcgtlccgga	994						
Db 777	CAGCGTCATCAACCCTGCTCAGAACACCCCGACAGCAGCAGCCGCCGCGCGCGACCCGTC	836						
QY 995	ggatattccagccggttgcggaagactc-----tcggcgcttctgcgttggttgcgga	1048						
Db 837	CTCGTGCCTCCGAGGCGGTGAGGAACCTGCTGCGCTCTCGGCGATCGCGAGCATGCGCGG	896						
QY 1049	tgagccgcatctccacacctccgattacagatttcatgtgcgtgcacatgaagaaagtgacaa	1108						
Db 897	CGGCGCGATCGCGCACCGCGAGCATGAGATGACGAGCAAGCTCATTCGCGCGCGGTGAAAG	956						
QY 1109	gattctgtactacgcagatgctgtctgcgtctgatagtacgcgcaaaacgcctccgatatca	1168						
Db 957	ACTGATCGCATCAACCTCCATCGCCACACCGGACGTTTCGGTTCGAGAACTCCGACCG	1016						
QY 1169	cgtagacttcagtcgcccacaaaggtttcacacacacactttgtccagcgacgcatctgtg	1228						
Db 1017	CCTCGATGTGACACCGCTGGGACAGCGCACCTCTCTCTTCCGGGTACGGGGTGCACCATG	1076						
QY 1229	ccttgcgacgacaccttgcgcgcgcgggaataatcgtcacctccacaaagaatgctgacac	1288						
Db 1077	CTGGGCGCAGAACCTGCGCCGCTCGAAGTCAAGTCACTCATCTCACCGTGTTCGACCG	1136						
QY 1289	gattctgcattcttcattgc 1310							
Db 1137	CATTCCGACCTGCGCTGCGC 1158							
RESULT 14								
LOCUS	STMCYP450	1274 bp	DNA	BCF	07-FEB-1999			
DEFINITION	Streptomyces carbophilus gene for cytochrome P-450sca-2, complete cds.							
ACCESSION	D30815							
VERSION	D30815.1	GI:1072316						
KEYWORDS	cytochrome P-450sca-2.							
SOURCE	Streptomyces carbophilus (isolate: SANK 62585) DNA.							
ORGANISM	Streptomyces carbophilus							
REFERENCE	Bacteria; Filicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycetaceae; Streptomycetes. 1 (bases 1 to 1274)							
AUTHORS	Serizawa, N.							
TITLE	Direct Submission							
JOURNAL	Submitted (25-MAY-1994) to the DDBJ/EMBL/Genbank databases.							
REFERENCE	Noufusa Serizawa, Sankyo Co., Ltd., Biomedical Research Laboratories; Hiromachi 1-2-58, Shinagawa, Tokyo 140, Japan (Tel:03-3492-3131(ex.3330), Fax:03-5436-8565)							
AUTHORS	2 (bases 1 to 1274)							
TITLE	Watanabe, I., Nara, F. and Serizawa, N.							
REFERENCE	Cloning, characterization and expression of the gene encoding cytochrome P-450sca-2 from Streptomyces carbophilus involved in production of pravastatin, a specific HMG-CoA reductase inhibitor							
AUTHORS	Gene 163 (1), 81-85 (1995)							
TITLE	96001248							
JOURNAL								
MEDLINE								
FEATURES	Location/Qualifiers							
SOURCE	1..1274							
	/organism="Streptomyces carbophilus"							
	/isolate="SANK 62585"							
	/db_xref="taxon:44059"							
	35..1267							
	/codon_start=1							
	/transl_table=1							
	/product="cytochrome P-450sca-2"							
	/protein_id="BA06492.1"							
	/db_xref="gi:1072317"							
	/translation="MTEMTKATFTFLTSOENAPAPADRTCPYQLTAYSRLEKDEPDLRPVLYGRRAMVYVTKHKAARLADLRSDRHLHAFPAISPEFKAFQSGPAFIMGDPHGRHRRMTISEFTVTKRKGMRPVERVHVFIDDMILAACTPALVDSOFAFPVSMTVCNHCVPAAHDEPFQDSKRYQAVAVDASNAAADDPERYIDGLITLESPTGCLGKLVTHQADGETIDRAELISTRLLLVGHETTASMTSLTITLLEHPDQAAKALADPSLVPAQVAVELLRLVLAIDADGRIATVADIIDGLGIEGVITVNSIANRDSVFE							

[illegible]



LOCUS PSECAMABA 1950 bp DNA BCT 01-FEB-2000  
DEFINITION Pseudomonas putida gene for cytochrome P-450cam, NADH-putidaredoxin  
reductase, putidaredoxin, partial cds and complete cds.  
ACCESSION D00528  
VERSION D00528.1 GI:216870  
KEYWORDS putidaredoxin; camC; NADH-putidaredoxin reductase; camA; camB;  
camphor; cytochrome P-450cam.  
SOURCE Pseudomonas putida (strain:ATCC17453 (strain PpG1)) DNA,  
clone:ppp45.  
ORGANISM Pseudomonas putida  
Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
Pseudomonas.  
REFERENCE 1 (bases 1 to 1950)  
AUTHORS Koga,H., Yamaguchi,E., Matsunaga,K., Aramaki,H. and Horiuchi,T.  
TITLE Cloning and nucleotide sequences of NADH-putidaredoxin reductase  
gene (camA) and putidaredoxin gene (camB) involved in cytochrome  
P-450cam hydroxylase of Pseudomonas putida  
J. Biochem. 106 (5), 831-836 (1989)  
JOURNAL 90130389  
MEDLINE  
COMMENT Submitted in computer readable form by H. Koga on 13-Jan-1990.  
FEATURES  
source  
1..1950  
location/Qualifiers  
/organism="Pseudomonas putida"  
/strain="ATCC17453 (strain PpG1)"  
/db\_xref="taxon:303"  
/clone="ppp45"  
/note="2 bp upstream of SmaI site.; total cellular DNA"  
1..93  
/gene="camC"  
<1..93  
/gene="camC"  
/note="carboxy terminus"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P-450cam"  
/protein\_id="BAA00412.1"  
/db\_xref="GI:216871"  
/translation="PGAQIQHKSIVGVQALPLVMDPATTKAV"  
101..104  
/note="SD sequence for camA"  
115..1383  
/gene="camA"  
115..1383  
/gene="camA"  
/gene="camA"  
/codon\_start=1  
/transl\_table=11  
/product="NADH-putidaredoxin reductase"  
/protein\_id="BAA00413.1"  
/db\_xref="GI:912426"  
/translation="MNNANDNVIVGTGLAGVEVAFGLRAGMEGNIRLVGDATVPIPH  
LPLSKAYLAGKATAESLYLRTPDAAQNIOLIGTQVATINRDRQVITLSDGRAID  
YDRLVATGGRPRPLPYASGAVKANNFRILRTLEDACIRROLIADNRLLVITGGYT  
GLEVAATAIKANHVTLLDIAARVLEVTAPVSAFVYEHLLHREGVDIRGTGCGFE  
MSTDQOKVAVLCEIDGTRLPLADLVIAIGLIPNCELSAAGLQVNDGIVINEHMQTSD  
PLIMAVGDCARFHSQLYDRWVRIEVPNALEQARKIAAIIICGKVPREDAPMPMSDOY  
EILKMWGSEGGDRIIVRGSIAQPDPSVFLQGDRLAVDTVNRPEFQNSKQIITD  
RLVEPNILGDESVPLKEIIAAKAKELSA"  
1427..1430  
/note="SD sequence for camB"  
1439..1762  
/gene="camB"  
1439..1762  
/gene="camB"  
/codon\_start=1  
/transl\_table=11  
/product="putidaredoxin"  
/protein\_id="BAA00414.1"  
/db\_xref="GI:216873"  
/translation="MSKVVYVSHDGTRELDVADGVSLMQAAVNGIYDIYDGGGSA  
SCATCHVYVNEAFTDKVPANEREICLBCVTALKPNSRLCCQIITPPLDGIYVDV  
PDRQW"  
1442..1759  
/gene="camB"

BASE COUNT 450 a 555 c 575 g 370 t  
ORIGIN  
Query Match 6.6%; Score 92; DB 1; Length 1950;  
Best local similarity 100.0%; Pred. No. 1.6e-09;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1311 ccgggtgcccaatcagcaagaagcgatcgtaagcggtgcaagcactccctctg 1370  
|||||  
Db 1 CCGGGTGCCTCGATTTCAGACAGAGCGGCGATCGTCAGCGCGCTGCAAGCACTCCCTCTG 60  
Oy 1371 gctcgatccggcgactaccacaagcgatata 1402  
|||||  
Db 61 GTCTGGGATCCGGCGACTACCAAGCGGTATTA 92

Search completed: October 4, 2000, 06:41:30  
Job time: 13125 sec

mat\_peptide



GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 03:16:31 ; Search time 101.52 Seconds

(without alignments)  
3455.174 Million cell updates/sec

Title: US-09-246-451-1

Perfect score: 1402  
Sequence: 1 ctgcagatcgtatccgct.....ggcactaccagaacggtatata 1402

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq-36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1402	100.0	1578	X19926	Pseudomonas putida
2	652.2	46.5	1356	X19916	Bacterial and mamm
3	140.8	10.0	110000	V30458_4	Continuation (5 of
4	140.8	10.0	110000	V30459_4	Continuation (5 of
5	113	8.1	15079	Q91580	S. clavuligerus cl
6	110.6	7.9	5821	T58686	DNA encoding S. fr
7	97.6	7.0	1233	O61452	Cytochrome P450 SC
8	88.2	6.3	1879	Q11126	Sequence encoding
9	83.4	5.9	1194	T58555	Streptomycetes prist
10	83.4	5.5	4496	T58553	Streptomycetes prist
11	77.8	5.5	53789	V21187	Amycolatopsis medi
12	76.6	5.5	2168	Q73674	Mycinamycin IV hyd
13	74.6	5.3	1998	Q11127	Sequence encoding
14	73.6	5.2	1700	Q45569	Sequence soyC and
15	68.6	4.9	1400	Q14548	Eryf gene. New 6-d
16	58.8	4.2	6085	T70153	S. longisporoflavus
17	44.4	3.2	8169	V26609	Actinomadura hibis
18	40.4	2.9	35089	V27112	Adenovirus 17. Rec
19	39.8	2.8	503	Q14549	Eryf gene recognit
20	39	2.8	1239	N91576	Sequence of clone
21	39	2.8	1239	Q43436	Lol PI cDNA clone
22	38	2.7	1929	V57472	Sorghum bicolor (L
23	37.2	2.7	1000	V63929	Mycobacterium tube
24	37	2.6	4393	O81456	Orotidine-5'-monop
25	36.8	2.6	13987	T80415	Hybrid srmg/tylg O
26	36.8	2.6	43280	T80413	Tylosine synthase
27	36.6	2.6	1747	T12248	Cytochrome P450 is
28	36.4	2.6	8478	O81792	B. subtilis biotin
29	36.4	2.6	29879	Q46806	eryA region of S.
30	35.8	2.6	1164	O72603	Plant insecticidal
31	35.8	2.6	1286	V74187	Barley pathogen in
32	35.8	2.6	1286	X23274	H. vulgare cystein
33	35.8	2.6	3076	X23275	H. vulgare cystein

C	34	35.2	2.5	750	1	X32469	P. fluorescens mal
	35	35	2.5	1485	1	O87721	Human auxiliary cy
	36	35	2.5	1485	1	O87722	Human auxiliary cy
	37	35	2.5	1485	1	T17409	Human derived cyto
	38	35	2.5	1485	1	T28387	Human cytochrome p
	39	35	2.5	1485	1	T28388	Human cytochrome p
	40	34.6	2.5	9960	1	V58939	Mycobacterium smeg
	41	34.2	2.4	930	1	X25146	Wheat type I gluta
	42	34	2.4	716	1	V49535	harkiri cDNA. New
	43	34	2.4	954	1	O66592	GST-II-27 gene. PI
	44	34	2.4	4682	1	O34557	GST-II-27 cDNA. 27
	45	33.8	2.4	1564	1	O38955	Lipase gene #1. DN

ALIGNMENTS

RESULT	1
ID	X19926
AC	X19926; standard; DNA; 1578 BP.
DF	11-JUN-1999 (first entry)
DE	Pseudomonas putida cytochrome P450 protein P450-cam encoding DNA.
KW	Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;
KW	oxidase; hydrocarbon; carbon-hydrogen bond; hydroxylating;
KW	bioremediation; environmental pollutant; ss.
OS	Pseudomonas putida.
PN	W09908812-A1.
PD	25-FEB-1999.
PF	17-AUG-1998; U16979.
PR	20-AUG-1997; US-056754.
PA	(UVRP) UNIV ROCHESTER.
PI	Jones JP, Shimoi M;
DR	WPI: 99-190131/16.
DR	P-PSDB: Y04128.
PT	New P450 fusion proteins - comprising a portion of a bacterial
PT	cytochrome P450 protein and a portion of a mammalian cytochrome P450
PT	protein
PS	disclosure; Page 13-14; 51pp; English.
CC	The present invention describes a fusion proteins comprising a portion
CC	of a bacterial cytochrome P450 protein and also a portion of a mammalian
CC	cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or
CC	any compound having a carbon-hydrogen bond. The fusion protein can be
CC	used for hydroxylating a compound to be oxidised. It can also be used in
CC	the bioremediation of an environmental pollutant. Since the fusion
CC	protein is soluble, it can be subject to structural elucidation by X-ray
CC	crystallography for designing functional proteins. It can be readily
CC	expressed in soil bacteria to facilitate bioremediation. The present
CC	sequence encodes Pseudomonas putida cytochrome P450 protein P450-cam
CC	from the present invention.
CC	Sequence 1578 BP; 350 A; 490 C; 436 G; 302 T;
SO	
Query Match	100.0%; Score 1402; DB 1; Length 1578;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1402;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 ctgcagatcgtatccgctgagcagatcgtatccacacggttttcacatgacgagcc 60
DB	1 CTGCAGATCGTATCCGCTGAGCAGATCGATCAACCGCTTTTCATCGACGAGGCC 60
OY	61 agcaagacactgaactggtcaagagcagagcactgataaaccccgatgactcact 120
DB	61 AGCAAGACACTGAACGTGTAAAGCAGACACTGATCAACCGGTGATCGACTCCACT 120
OY	121 cttagcacaaccgcggttcacagagaaacaacaatgacgactgataaacacataaagca 180
DB	121 CTTTAGCAACCCGCGTTCACGAGAAACAACAATGACGACGTGAACCACTAACAAAGCA 180
OY	181 agcgcaatctgcccctctgcacacccatgtgcagagcactggtatctgcagactgaca 240
DB	181 AGCGCAATCTTGCCCTCTGTCACCCCATGTGCGAGAGACACTGTGATTCGACTTCGACA 240

QY 241 tglacaatccgctgaatctgtctgcgagctgcaggaagccttgcgaagatcctgcgaagat 300  
DB 241 TGTAAATCCGCTGGAATCTGTCCGGCTGCAGAGAGCCCTGGGCGATTCTCAAGAAAT 300  
QY 301 caaagctaccgagctctgtgtgtgagctcgtcgaacgagcggaactgtgatacctgcg 360  
DB 301 CAAAGCTACCGGATCTGTGTGAGCTGCTGCAACGGCGGACACTGATCGCACCTCGCG 360  
QY 361 gccactgtatcctgtgagcctcgaagattacacgcaactttccagcgagctgcgcgttca 420  
DB 361 GCCAACTGATCCCTGAGGCGCTATGAAGATTACGGCCACTTTTCAGAGATGCGCCGTTC 420  
QY 421 tccctcgtgaagcggcggaagcctacgactlccaccactcgaatgagatcgcgcccagc 480  
DB 421 TCCCTGTGTAAGCGCGGGAAGCCTACGACTTTCACCTCCAGCTGATGCGCCGAGC 480  
QY 481 aggcgcagttctgtgctcgtgcgaacaaagtgtgtgacatgcggtgtgtgtgataagctg 540  
DB 481 AGGCGCAGTTCTGTGGCTGGCCACCAAGTGTGTGCAATGCGGTGTGATTAAGCTGG 540  
QY 541 agaacggatccaggaagctgctgtcgtctgatacgaagcctgcgcccgaaggaagat 600  
DB 541 AGAACCGGATCCAGAGACTGCTGCTGCTGATCGAGAGCTTGCCTCCCAAGAGACT 600  
QY 601 gcaacttcacggaagctacgacgcaacccctcccgatcagcatcttcattgtctgcag 660  
DB 601 GCAACTTTCACCGAGACTACGCCGAACCCCTCCGATACCGATCTTCATCTGCTGCGAG 660  
QY 661 gtctaccggaagaagatatccgcgaacttgaataactaaagatcagatgacccgtccg 720  
DB 661 GTCTACCGGAAGAATATCCCGCACTTGAATACCTTAACGGATGATACCGCTCCGG 720  
QY 721 atgagcaatgactctgcgaaggaagagcgctcgaactctcgaactctcgaactatca 780  
DB 721 ATGAGCAATGACTCTGCGAGAGGCCAAGAGGCGCTCTACGATATCTATATCCGATCA 780  
QY 781 tcgaagaacgcaagcagaagccggaagcagcgtalcaagcatcgtltgcgaagcag 840  
DB 781 TCGAGCAACGCAAGCAAGAACCGGGAACCGACTATACCATCTGTTGCCAACGGCCAGG 840  
QY 841 tcaatggggcagccgataccagctgaacgaagcaagagatgtgtgtgtcttactgtcg 900  
DB 841 TCAATGGGGCAGCCGATCACCAAGTACGAAACCAAGAGAGTGTGGCTGTACTGTGTCG 900  
QY 901 ggcgctgtgatacgtgtgtaactttctcgaactcagcatcagtgaggttccctgcgaaga 960  
DB 901 GCGGCTGTGATACGCTGTGTAATTTCTCTCAAGCTTCAAGATGGAGTTCTTGCCAAAAAGCC 960  
QY 961 cggagcatcgcaggaagctgtatcgaagcgtlcccgagcgtaltccagcgcgttgcgaaga 1020  
DB 961 CGGAGCATCGCCAGAGCTGATGAGCGTCCGAGCGTATTTCCAGCGGTTGGGAGGAAAC 1020  
QY 1021 tactcggcgctctcgtcgtgtgtgcgaatgcgcgcatcctcaactcctcgaatgaagttc 1080  
DB 1021 TACTCGGCGCTCTCTGCTGTGTCGATGCGATGCGCATCTCCACACTCCGATTAAGAGTTTC 1080  
QY 1081 atgagctgcaactggaagaagagtgacacagatcctgtcaaccgagatgagctgtgcgagct 1140  
DB 1081 ATGAGCTGCAACTGGAAGAAAGGTGACCAAGATCTGCTACCGCGAGATGCTGTGCGCTGG 1140  
QY 1141 atgagcgcgaaaaagcgtcgcgaatgcaagctgacttcagtcgcgaagaagtttccaca 1200  
DB 1141 ATGAGCGCGCAAAAGCCTGCGCATGACATGCACTTTCAGTCCCAAAAGTTCACACA 1200  
QY 1201 ccaactttggcagcagcagcagcatctgtgtccttggcagcagcagcagcagcaggaatca 1260  
DB 1201 CCAACTTTGGCCAGCGCAGCATCTGTGCTTGGCAGCACTGTGCGCGCGGGAATCA 1260  
QY 1261 tcgtcaacctcaagagatgctgacaggaatctcgaactctcattgcccgggtgtgcc 1320  
DB 1261 TCGTCAACCTCAAGAAATGCTGACCAAGATTCTTCACTTTGATTTGCCCGGATGCC 1320  
QY 1321 agattcagacaagaagcgcgcatcgtcagcggtgcagagcactccctcgtgtcggatc 1380

DB 1321 AGATTGAGCAACAAGACCGGATGTCAGCGGCTGAGGACGATCCCTGTGCGGATC 1380  
QY 1381 cgggactaccacaagcagatata 1402  
DB 1381 CGGCGACTACCAAAAGCGGATATA 1402  
  
RESULT 2  
X19916  
ID X19916 standard; DNA; 1356 BP.  
AC X19916;  
DT 11-JUN-1999 (first entry)  
DE Bacterial and mammalian chimeric cytochrome P450 protein encoding DNA.  
KW Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;  
KW oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;  
KW bioremediation; environmental pollutant; ss.  
OS Synthetic.  
PN MO9908812-A1.  
PD 25-FEB-1999.  
PE 17-AUG-1998; U16979.  
PR 20-AUG-1997; US-056754.  
PA (U9RP ) UNIV ROCHESTER.  
PI Jones JP, Shimoi M;  
DR MPI: 99-190131/16.  
P-PSDB: Y04126.  
PT New P450 fusion proteins - comprising a portion of a bacterial  
PT cytochrome P450 protein and a portion of a mammalian cytochrome P450  
PT protein  
PS Claim 10; Page 5-6; 51pp; English.  
CC The present sequence encodes a fusion proteins comprising a portion of a  
CC bacterial cytochrome P450 protein and also a portion of a mammalian  
CC cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or  
CC any compound having a carbon-hydrogen bond. The fusion protein can be  
CC used for hydroxylating a carbon compound to be oxidised. It can also be used in  
CC the bioremediation of an environmental pollutant. Since the fusion  
CC protein is soluble, it can be subject to structural elucidation by X-ray  
CC crystallography for designing functional proteins. It can be readily  
CC expressed in soil bacteria to facilitate bioremediation.  
SQ Sequence 1356 BP; 349 A; 386 C; 323 G; 298 T;  
  
Query Match 46.5%; Score 652.2; DB 1; Length 1356;  
Best Local Similarity 99.5%; Pred. No. 2e-163;  
Matches 654; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 156 atgagcactgaaacacatacaagaacgcaactgtgcctctgcacccaatgtgca 215  
DB 1 ATGAGCACTGAAACCATACAAAGCCCAATCTTGCCTTGCACCCCATGTGCA 60  
QY 216 gaagcactgattgactcgaactcgaatgaacacgtgaaatctgtcgcggcgtag 275  
DB 61 GAGCACCCTGGTATTTCACATTTCGACATGTACAAATCCGTCAATTTGTTGCGGCGTGCAG 120  
QY 276 gaagccttggcagttctgcgaagaatacaaaagctacccgagatcgtgtgcagcgtcgaac 335  
DB 121 GAGGCTGGGCACTTGTGCAAGAATCAAAACGTAACGATCTGTGTGAGCTGCTGCAAC 180  
QY 336 ggcgcgaactgtatgcgacactcgcgcgcaactgatactcgtgagagcctatgaagattaccgc 395  
DB 181 GGGGGAACACTGAGTCCCACTGCGGCGCAACTGATCCGTGAGGCGCTATGAAGATTACCGC 240  
QY 396 caatttcagcagagtgccggttcatccctcgtgaaagcggcggaagcctcagactcatt 455  
DB 241 CACTTTTCCAGCAGTAGTCCCGTTTCATTCCTCGTGAAGCCGCGCAACCTTACATTT 300  
QY 456 cccacactgatgatccgcgcgagcagcgcagttcgtgtgcgtgcgcgaaccaagtgtt 515  
DB 301 CCAACCTGATGATGCCGCCGCGAGCGGCACTTTCGTGCGTGGCCCAACGAAGGT 360  
QY 516 ggcatacgcggtgtgtgataagctggaagcagatcccaagagactgctgcgtgatac 575  
DB 361 GGATATCGCGGTGTGATTAAGCTGGAACCGGATCCAGAGAGCTGCGCTGCTGATATC 420

```
Oy 576 gaagactgccccgaagacatgcaactcaaccaggactacgcgcgaacctccg 635
    |||||||
Db 421 GAGAGCTCGGCCGCCGCAAGACACTGCACTTACCGAGACACTACCGCAACCTTCCG 480
Oy 636 atacgactcctcgtcgtcgcagatctacccggaagaatalcccgacttgaatlac 695
    |||||||
Db 481 ATAGCATCTTCAATGCTGCTCGCAGGCTACCGGAGAGATATCCCGCACTTGAAATAC 540
Oy 696 ctacagatcatagatgacccgtccggtatgcagcatgaccccttcgacagccaaaggcg 755
    |||||||
Db 541 CTACAGCATACAGTACACCCCTCCGAGATGACACTTCCGAGAGGCCAAGAGGCG 600
Oy 756 ctctacgactcctgataccatcatcagcagcagcagcagcagcagcagcagcagcagc 812
    |||||||
Db 601 CTCTACGACTATCTGATACCGATCATCGACCAACGAGGAGAGCCGGAGATGAAAC 657
```

## RESULT 3

Continuation (5 of 6) of V30458 from base 400001 (Rhizobium species plasmid pNGR234a.)  
WP Sequence split into 6 fragments LOCUS V30458 Accession V30458

Fragment Name	Begin	End
WP V30458_0	1	110000
WP V30458_1	100001	210000
WP V30458_2	200001	310000
WP V30458_3	300001	410000
WP V30458_4	400001	510000
WP V30458_5	500001	534720

Query Match 10.0%; Score 140.8; DB 1; Length 110000;

Best Local Similarity 48.5%; Pred. No. 1.6e-27; Matches 478; Conservative 0; Mismatches 502; Indels 6; Gaps 3;

```
Oy 324 actcgtgcgaacgagcagacatgcatcgcacatgcgcgcgaactgcatcgltgagccat 383
    |||||
Db 46785 ACGCGCATGAGAGAGGTACTGGGTATACACAGGCGAGACACGCGCGGTCCTT 46844
Oy 384 gaagattacccgaactcttccacgagatgcgcgttccatccctcgtgtaagccgcggaagc 443
    |||||
Db 46845 GAGGATACAGAGACCTTTTCCACGACATCTTCCCTCGGACATGCGCACTTGGGAACAC 46904
Oy 444 taagactcattcccaactcgaatgcatcgcgcgcgaagcagcagcagcagcagcagcagc 503
    |||||
Db 46905 TGCGCGGTCTCCCTCGAATTGACCCCGCCCATGAGTGTTCGTGACATCTCTA 46964
Oy 504 aaccaaagtgtgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 563
    |||||
Db 46965 AATCCTCTGTTTCATCCAGCGGGGTGTGGCGTGAAGCGCACTATCCACAGAGCA 47024
Oy 564 tgcctcgtatcgagagcct---gcgcgcgaagagacatgcaactcaccagagactac 620
    |||||||
Db 47025 GGGCGGTGATGACATGTCATGCGAAGAGAACACAGCTGTGACGTGATGAGAGATTTTC 47084
Oy 621 ggcgaacctcccgatagcagcatctcattgctcgcaggtctaacggaagaagatc 680
    |||||
Db 47085 GCGTTGCGCTTTCATGCTTTCATGCTTTCATGCTTTCATGCTTTCATGCTTTCATGCTTTC 47143
Oy 681 ccgacttgaataactcaagacatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 740
    |||||
Db 47144 CGAAGTACTTGTGGGTGGGTAAAGCATTTGCTGACAGGACGACGACGACGACGACGACGACG 47201
Oy 741 gaagccaagaagagcgtctacgactatcgtatccgactcagacatcagacagcagcagcagcag 800
    |||||
Db 47202 GCAGCAGCCCGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 47261
Oy 801 ccggaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 860
    |||||
Db 47262 CCAAGCTTCATTTCAAGACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 47321
Oy 861 agtgaacaaagcagaagatgtgtgctgttactggttcgagcgcgtggaatacgtgtgtc 920
    |||||
```

## RESULT 4

Continuation (5 of 6) of V30459 from base 400001 (Rhizobium species symbiotic plasmid  
WP Sequence split into 6 fragments LOCUS V30459 Accession V30459

Fragment Name	Begin	End
WP V30459_0	1	110000
WP V30459_1	100001	210000
WP V30459_2	200001	310000
WP V30459_3	300001	410000
WP V30459_4	400001	510000
WP V30459_5	500001	536165

Query Match 10.0%; Score 140.8; DB 1; Length 110000;

Best Local Similarity 48.5%; Pred. No. 1.6e-27; Matches 478; Conservative 0; Mismatches 502; Indels 6; Gaps 3;

```
Oy 324 actcgtgcgaacgagcagacatgcatcgcacatgcgcgcgaactgcatcgltgagccat 383
    |||||
Db 46785 ACGCGCATGAGAGAGGTACTGGGTATACACAGGCGAGACACGCGCGGTCCTT 46844
Oy 384 gaagattacccgaactcttccacgagatgcgcgttccatccctcgtgtaagccgcggaagc 443
    |||||
Db 46845 GAGGATACAGAGACCTTTTCCACGACATCTTGGCTGCGCACTTGGGAACAC 46904
Oy 444 taagactcattcccaactcgaatgcatcgcgcgcgaagcagcagcagcagcagcagcagcagc 503
    |||||
Db 46905 TGCGCGGTCTCCCTCGAATTGACCCCGCCCATGAGTGTTCGTGACATCTCTA 46964
Oy 504 aaccaaagtgtgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 563
    |||||
Db 46965 AATCCTCTGTTTCATCCAGCGGGGTGTGGCGTGAAGCGCACTATCCACAGAGCA 47024
Oy 564 tgcctcgtatcgagagcct---gcgcgcgaagagacatgcaactcaccagagactac 620
    |||||||
Db 47025 GGGCGGTGATGACATGTCATGCGAAGAGAACACAGCTGTGACGTGATGAGAGATTTTC 47084
Oy 621 ggcgaacctcccgatagcagcatctcattgctcgcaggtctaacggaagaagatc 680
    |||||
Db 47085 GCGTTGCGCTTTCATGCTTTCATGCTTTCATGCTTTCATGCTTTCATGCTTTCATGCTTTC 47143
Oy 681 ccgacttgaataactcaagacatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 740
    |||||
```

```
Db 47144 CGAAGTACTTGTGGCTGGGTAAAGCGATTGTGCTCCAGCGCAACGAGAAAGCGAAG-- 47201
Oy 741 gaagccaagaagagcgctctacgactatcgtatccgatactgagcaagcgaagcaag 800
Db 47202 GCAACACACCCCGCTGGCTGGCTGCTTTATTTAGCAAAATGCGAGGATGCGCGCAACTCG 47261
Oy 801 ccgggaacgcagcgtatcagcatcgtgtgcacaacgcgcaaggltcaatggcgacgcatacc 860
Db 47262 CCAAGCTGTGATTTGATGACCTTCTGCTTCAGCGCAAAATGAGGGCGCTCTTAAACA 47321
Oy 861 aatgaagcaagcaagagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 920
Db 47322 GAAGAGAGAAATCGCGCGGCAATCGGTGCTTTCTTTCTGCTCGCGGGGCTCGACACGTTGGCC 47381
Oy 921 aatttcacagcttcagcagagagcttcaggcgaagaagcccgagcaatcgccgagagctg 980
Db 47382 GCCGCGCTTTGGCTTTGACATAGGCTTATCTTGGACAGCAATCCAAAGCACCGAGAGTTCTTG 47441
Oy 981 atcgagcgctccgagcgatctccagccgcttgagagaaactactcggcgcttcgcgtg 1040
Db 47442 CGAAGACGACCGGCTGGCTGGGCTGGCGCTGAGGAACTGCTGGCGGCTATTTCAACC 47501
Oy 1041 gtgcgcgagtcgcgcacatcctcactccgattacagagttcatgtgcgtgcaactgaagaa 1100
Db 47502 GTTCAGATTAATCCGCGTGGCGGCGCAAGACATGAGTTCCGAAGCGCTGCTATCCGTGAG 47561
Oy 1101 ggtgacacagatcctgcacccgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1160
Db 47562 GGTGATTTATGTTTCTTCTTCCCGCATGATGCTAATGAGATCCGCTGTAATTTAACTTG 47621
Oy 1161 ccgagtcagctgcagctcagctcagctcagctcagctcagctcagctcagctcagctcagct 1220
Db 47622 CCGAAGCACTATTCGATTCGGGACGACGACGATACGACGACGCGCTTTGGCTATGAGCA 47681
Oy 1221 catcgtgcctgtgcagcaactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1280
Db 47682 CACCTTTGCCACGAGCGCATCTTCGACGCGGAGAAATCGTCAATGCGCTTACGATGAG 47741
Oy 1281 ctgacccagatctcagctcagctcagctcagctcagctcagctcagctcagctcagctcag 1306
Db 47742 TTAGCGCGCATCCACCTTCCGAT 47767

RESULT 5
ID 091580 standard; DNA; 15079 BP.
AC 091580;
DE 13-NOV-1995 (first entry)
DE S. clavuligerus cla gene region.
KW clavulanic acid; clavulinate; antibiotic; beta-lactamase-inhibitor;
OS Streptomyces clavuligerus.
FH key location/Qualifiers
FT cds complement (49..1745)
FT FT /tag= a
FT FT /label= ORF1
FT FT 2216..3937
FT FT /tag= b
FT FT /label= ORF2
FT FT 3940..5481
FT FT /tag= c
FT FT /label= ORF3
FT FT 5654..6595
FT FT /tag= d
FT FT /label= ORF4
FT FT /note= "cla gene"
FT FT 6611..7588
FT FT /tag= e
FT FT /label= ORF5
FT FT 7895..9076
FT FT /tag= f
FT FT /label= ORF6
```

```
FT FT cds 9241..10908
FT FT /tag= g
FT FT /label= ORF7
FT FT complement (10998..12296)
FT FT /tag= h
FT FT /label= ORF8
FT FT complement (12662..13365)
FT FT /tag= i
FT FT /label= ORF9
FT FT 13769..14995
FT FT /tag= j
FT FT /label= ORF10

CA2108113-A.
PN 09-APR-1995.
PF 08-OCT-1993; 108113.
PR 08-OCT-1993; CA-108113.
PA (UYAL-) UNIV ALBERTA.
PI Aldoc KA, Jensen SE, Paradkar AS;
DR WPI; 95-207301/28.
DR P-PSDB; R77858; R77859; R77860; R77861; R77862; R77863; R77864;
DR R77865; R77866; R77867.
PT clavulanic acid biosynthesis enzymes and corresp. DNA - useful for
PT biosynthesis of the antibiotic in Streptomyces hosts which do not
PT naturally produce clavulanate
PS disclosure; Fig.2; 41pp; English.
CC A library of S. clavuligerus NRRL 5741 genomic DNA fragments (15-22
CC kb) constructed in cosmid pLAFR3 was screened for the cla gene using
CC a probe based on a partial N-terminal sequence from the cla gene.
CC Isolated clone K611 included a 15kb fragment having the sequence
CC given in 091580 that included the cla gene (ORF4).
SQ Sequence 15079 BP; 2097 A; 5488 C; 5324 G; 2170 T;

Query Match 8.1%; Score 113; DB 1; Length 15079;
Best Local Similarity 51.3%; Pred. No. 1,9e-20;
Matches 291; Conservative 0; Mismatches 270; Indels 6; Gaps 1;

Oy 736 tcgcaagagccaagagcgctctacgactatctgataccgatactcagcaagcagc 795
Db 14340 TCGCCAAAGCGCCGGGACGACGACTGACGCGCTATCTCGGGAGCTGTGTGAGAGAGGATCG 14399
Oy 796 agaagccgggaacgcagcgtatcagcatcgtgtgcacaagcgagatgaatggcgagcga 855
Db 14400 AGAACC CGGCGACCGACCTGATGACCGCGCTGCTATCGACAGTGTGCGCGGCGATC 14459
Oy 856 tcaacagtgacggaacgaagaagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 915
Db 14460 TCGGGGTGAGAGATGTGCTCCATGTGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 14519
Oy 916 tggtaattcctcagctcagctcagctcagctcagctcagctcagctcagctcagctcagctcag 975
Db 14520 CCACAGGCCAGGGGAGCGCTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 14579
Oy 976 agctgatcagagcgctccgagcgatccagcgagcgcttcgaggaactcagcgagctctc 1035
Db 14580 GCCTTACCGAGGACCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 14639
Oy 1036 cgctgtgtgcagtgccgagctcagctcagctcagctcagctcagctcagctcagctcagctcag 1089
Db 14640 CCAATCGTGCAGAGCGGCTGCGCCGCTGCGCGGTGAGGACGTCCAGTCCAGATGTGC 14699
Oy 1090 aactgaagaagtgacagcagatcctcagcagagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1149
Db 14700 TCAATCGGGGGGGGCGAGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 14759
Oy 1150 aaaaagcgtgcccagatgcagctcagctcagctcagctcagctcagctcagctcagctcagctcag 1209
Db 14760 TCTTCCCGGACCGGACCGGCTGAGCTGTGACCGGACCGGCGCCGCAATCTGCGCTTGC 14819
Oy 1210 gccagcgcaagcactcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1269
Db 14820 GCCACGGCATGACACAGTGTGCTGTGGCCAGTGTGCGCGGGGTGTGTGTGTGTGTGTGTGTGTGT 14879
```

QY	1270	tcaagaatgctgaccagattcctg	1296
Db	14880	TCGCCGCGGTCTGCCTGGATGCCG	14906

RESULT	6
T58686/c	
ID	T58686 standard; DNA; 5821 BP.

Result	6
TS8686/C	TS8686 standard; DNA: 5821 BP.
AC	TS8686;
DT	17-FEB-1998 (first entry)
DE	DNA encoding S. fradiae tylosin biosynthesis gene products.
KW	tylosin; biosynthetic gene product; production; antibiotic; ds.
OS	Streptomyces fradiae.
FH	Key
FT	CDS
FT	location/Qualifiers
FT	/*tag= a
FT	/product= tlyE
FT	complement (1277. .1957)
FT	/*tag= b
FT	/product= tlyD
FT	complement (2286. .4049)
FT	/*tag= c
FT	/product= tlyH
FT	4094. .4927
FT	/*tag= d
FT	/product= tlyF
FT	complement (5046. .5660)
FT	/*tag= e
FT	/product= tlyI
PN	US5672497-A.
PD	30-SEP-1997.
PE	21-DEC-1995; 575843.
PR	12-MAY-1989; US-351350.
PR	21-MAR-1986; US-842330.
PR	25-JUL-1986; US-890670.
PR	24-FEB-1987; US-018937.
PR	06-AUG-1991; US-742222.
PR	28-JUL-1993; US-107732.
PR	17-FEB-1994; US-198672.
PR	21-DEC-1995; US-575843.
PA	(EPLI.) LILLY & CO ELI.
PI	Cox KL, Fishman SE, Hershberger CL, Seno ET;
PI	WPI: 97-48860/45.
PI	P-P-SDB: W33272, W33273, W33274, W33275, W33276.
PT	DNA encoding Streptomyces fradiae tylosin biosynthesis gene products
PT	- for increasing tylosin production in Streptomyces spp.
PS	Claim 1; Columns 21-38; 38pp; English.
CC	The present sequence encodes Streptomyces fradiae tylosin
CC	biosynthetic gene products, useful to increase the production of
CC	the antibiotic tylosin in Streptomyces spp.
CC	Sequence 5821 BP; 799 A; 2066 C; 2087 G; 863 T;

Query Match	7.9%	Score 110.6	DB: 1	Length 5821
Best Local Similarity	47.4%	Pred. No. 6.1e-20		
Matches 397	Conservative 0	Mismatches 434	Indels 6	Gaps 2
<hr/>				
QY	466	tgaatccgcgcccagcagcgccagttctgtgctgtgccaaccaagtggltgycalgcg	525	
Db	3493	TGGACCCGCCCGACACACGAGGCGTCCCGGCGCATTCATCCCGAGTTCGGCGCGGC	3434	
<hr/>				
QY	526	tgtgtgataaagcttgagaacccggtatccagagagctgtgctgtctgtatcgagagctgc	585	
Db	3433	GGGTGCGGGACTCTGGCCCTTCCTCGTGAACAGATGTGACCGGCTGTGTGATGACCTCA	3374	
<hr/>				
QY	586	gcccgcaag---acagtgcacacttcacacggaagctacgcgcgaacccttccgatacgca	642	
Db	3373	CCGCGCCCGGGGACAGAGCGGAGTGTGCTGTGGCGCATTTGGCGCTCCCAATGGCGACCCAG	3314	
<hr/>				
QY	643	tcttcatagtcgtctcgcaaggtctaccggaagaagatataccgcacttgaataactcaagc	702	
Db	3313	TGATCTGGCGCGCTGTCCACATTCCTCTTCAGAGACCGGAGTACTACTCCAGAGCCGACCG	3254	

QY	703	atcagatgaaccgctccgagatgagcaacttgacacttcgagagagccaaagagcgctctacg	762
Db	3253	AACAGGCGACCCCGCCCGCGCCCGGACGAGCGCGGTGGAGCGCGCTGTGGAGCTGCCG	3194
QY	763	actatctgataccgatatcatcgagaacagcgagcgcaagaagccgggaaccgagctatcagca	822
Db	3193	ACTTACTTGCACCGGCTGATCTAGCCGGCAGACCGGCCGGGAAATCCGGCGACGGGAAATCTCG	3134
QY	823	tcgttgcacaaagcgccagatcaatctgagcgacgatacacaagatgacgaagcacaagagatgt	882
Db	3123	GCAGAGATGTTGGCGCAGGCGCCCGCGGTGGGGGCTGTGTGCACGCGGCACGTCTGGACACAG	3074
QY	883	gttgccctgtatcctggtcgcgcgcctgtatacgtgtgtcatttcctcagcttcaacatg	942
Db	3073	CGGTGCTCTCTCTGCGCGCGGCGACAGACACAGCGCCAGCATGTCTCACGATGAGCTGC	3014
QY	943	agttctctgcacaaagcccgagagcatctgcaggaagctctgagagcgctcccgagcgatlc	1002
Db	3013	TGCTTCTCTTACACACACCCACAGGCGCTGGCGCCAACTCACCGGAAATCCCGCGCTGTTCG	2954
QY	1003	cagccgcttgcgaagaactactcgcgcgtcttcgcgtgtgttcgcgata--gcgcatacc	1055
Db	2953	CGGGCGCGGTGGACGAACTGCTGGCGTATGTGTGCATCGCGCAGACGGCGTGGCGCGTCCG	2894
QY	1060	tcacctccgattacgagttcatatgctgctgcaactgaagaagtgtaaccagatcctgtctac	1113
Db	2893	CCACGCGCGAATCATGAGATCTACACGGCCACACATTCGCGCGCGCGCGCTGTGTCTTCC	2834
QY	1120	cgcgaatgtctctgycgcgtgagatgagcggaaaagcgctgcgcgaatgcagcagcgtcgactta	1177
Db	2833	TGCTTGGCCGCGGCCAACCGGGAGACAGGCGCCCTTCTTCCGAACCGGAGCGCTTTCGAAATCC	2774
QY	1180	gtcgcccaaaaggtlctcacacaaccaaccttggccaagcagacacatctgtgtccttggccagc	1233
Db	2773	ACCGGTCCGCGAGCGGCGAATGTGCTTGGGCTACGGACCCACGACGATGCTGTGGGACAA	2714
QY	1240	acctgtgcccgcgggaaatcatcgttcaacctcaagaagtgtgtgcacagatattctgt	1296
Db	2713	ACCTGCGCGGATGGAATGTGAATGTGCGCTCGCGCGCGGTGTGGAGAGAGCTGTCCG	2657

RESULT	7	
ID	061452	
AC	061452 standard; DNA; 1233 BP.	
DT	24-NOV-1994 (first entry)	
DE	Cytochrome P450 SCA-2 gene.	
KW	Cytochrome P450 SCA-2; Streptomyces carboxyllus;	
OS	Treatment of hyperlipidemia; drug preparation; ss.	
FT	Streptomyces carboxyllus.	
FT	Key	Location/Qualifiers
FT	cds	1..1233
FT		/*tag= a
FT		/product= cytochrome P-450 containing activity
PM	J06070780-A.	
PD	15-MAR-1994.	
PF	28-AUG-1992; 229969.	
PR	28-AUG-1992; JP-229969.	
PA	(SANKY ) SANKYO CO LTD.	
DR	WPI; 94-128679/16.	
DR	P-PSDB; R51368.	
PT	Cytochrome P450 SCA-2 gene - from Streptomyces carboxyllus,	
PT	useful in treatment of hyperlipidemia	
PS	Claim 3; Page 12-14; 18pp; Japanese.	
CC	061452 encodes a protein having cytochrome P-450 activity	
CC	SCA-2 can be prepared commercially for use in the preparation of	
CC	drug for the treatment of hyperlipidemia.	
SO	Sequence 1233 BP; 207 A; 472 C; 372 G; 182 T;	





```
QY 1057 lctccacccgattacagattcatggtcgctgcaactgaagaaggtgacagatccctgc 1116
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1020 TCGCCACGCGGACATCGAGTTCGAGGGGACATCATCGGGGCGGCGAGGGCGTGATCG 1079
QY 1117 tacgcagatgcgtcgtcgtcgtgtagcgagaaacgcgtcccgatgacgtgcagct 1176
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1080 TCGTCMACTCGATGACCCACGGGACGGCAGCGTGTACGAGGAGCCGGACGCCCTCCGACA 1139
QY 1177 tcagtcgcgaagaaggttcaacacaccccttgccacagcgagcatctgtcgttgcc 1236
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1140 TCGACCGTCCGCGCGCCACCATCTCGCTTCGCGTTCGGCGTGCACCAAGTCCGGGCC 1199
QY 1237 agcacctggccgcgcggaatacatcgtcacccccaag 1275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1200 AGAACCTCGCCCGCGCTGAGCTGGAGGTCACTCTCAACG 1238

RESULT 9
ID T58555 standard; cDNA; 1194 BP.
AC T58555:
DF 02-APR-1997 (first entry)
DE Streptomyces pristinaespiralis snbF gene.
KW Streptogramin B; antibiotic; biosynthesis; pristinamycin;
KM virginiamycin; piperolic acid; cyclodeamination; papa; snbA; snbF;
PI p1pA; 3-hydroxypicolinic acid; hydroxylation; ds.
OS Streptomyces pristinaespiralis.
FH Key Location/Qualifiers
FT mat_peptide 1..1194
FT /*tag= a
FT /product= snbF
PD MO9601901-A1.
PN 25-JAN-1996.
PE 04-JUL-1995; F00889.
PR 08-JUL-1994; FR-008478.
PA (RHON ) RHONE POULENC RORER SA.
PI Barriere JC, Blanc V, Bianche F, Crouzet J;
PI Debussche L, Paris JM, Thibaut D, Bamas-Jacques N;
PI Dutruc-Rosset G, Famechon A;
PI WPI: 96-097631/10.
DR P-PSDB; W11585.
PT New streptogramin B deriva. useful as antibiotics - produced by new
PT mutants of Streptomyces having altered genes for streptogramin B
PT biosynthesis
PS Claim 18: Page 113-114; 146pp: French.
CC The papa gene of S.pristinaespiralis is involved in the biosynthesis
CC of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for
CC pristinamycin 1A. Upstream of the papa gene, on the complementary
CC strand, is the snbA gene coding for 3-hydroxypicolinic acid-AMP ligase.
CC The region between these two genes was sequenced and two open reading
CC frames were identified. The first (p1pA) decodes to an amino acid
CC sequence with homology to ornithine cyclodeaminase from Agrobacterium
CC tumefaciens. The p1pA gene product is likely to catalyse the cyclo-
CC deamination of lysine, leading to production of piperolic acid.
CC Mutations in the p1pA gene were shown to affect piperolic acid
CC synthesis but not the synthesis of 3-hydroxypicolinic acid. The second
CC open reading frame (snbF) could be decoded to give a product with
CC homology to hydroxylases of the cytochrome P450 type.
CC Disruption of the p1pA and snbF genes can be used to produce
CC strains of S.pristinaespiralis which are unable to produce the
CC antibiotic pristinamycin I but which may be able to produce new,
CC modified forms of it.
SQ Sequence 1194 BP; 189 A; 522 C; 321 G; 162 T;
```

```
Query Match 5.9%; Score 83.4; DB 1; Length 1194;
Best Local Similarity 45.7%; Pred. No. 5,7e-13;
Matches 291; Conservative 0; Mismatches 346; Indels 0; Gaps 0;
```

```
QY 734 ctgcgcagagcgaagggcgctctacgactatcgtatccagatcatcgagcaagcgag 793
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 543 CGTCMAAGACCATGCGCGCATGCAACGCTACTCTCACGACCAACATCACCGCGCG 602
```

```
QY 794 gcagaagccgggaaccgacgcatatcagatcgtltgccaagcgaggtcaatggcgagc 853
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 603 CGCCCGCCCGCGGAGACGACATGATCTCCGACACTGTCCGCCCGCGTGAGGGGAGAACG 662
QY 854 gataccagatgaagcaagcgaagagatgtgtgcctgttactgtgtcggcgagctgatatc 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 663 ACTCACCGACGAGCAGATGTCGTAATTCGGGGCGCTGCTGATGAGGCCGACAGTCTC 722
QY 914 ggtgtcaattctctcagatcagatgagttctctgcgcaaaagcccgagcatcgcca 973
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 723 CACTCTCATGCTGTCCGCAACACCGTCTGTGCTGAAGAGCAACCCCGGGCGAGGC 782
QY 974 ggaagtcgtagcagagctcccgagcgatattcacagcgcttgaggaagaaactccggcgtt 1033
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 783 CGCCGCGCCCGCCGACCGGTCCTGATCCCGCTGTATTCGAAGAAGTACTGGGCTGCG 842
QY 1034 ctgcgtgtgtgcagatgagcgcacatccatccatccgattacagattcattggtcgtaact 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 843 GCCGCGGATCAACCGTCAATGAGCCCGGTCACCAACGACCGTCTGCGCGGACAC 902
QY 1094 gaagaaggtgacagatccctgcgtacgcgagatgtgtgtcgtgtgtagtgagcgcaaaa 1133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 903 CATCCCGCCGAGCGCATGTCGTGCTCCCTGCTGTCCGCCAACACGACGACAGT 962
QY 1154 cgctcgccgattgacgctgcgacttcagtcgcaaaagttcaacaccactttggcca 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 963 CTTACACGACCCCGACACCTCTGACCTCGCCCGGAGAGCCCGCAGATCGCTTCGCGCA 1022
QY 1214 cgcagaccatctgtgccttggcagcaacttgcgcgcgcggaagaaatcagtcacacctcaa 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1023 CGGCATCCACTACTGCTCGGGCGCCCGCTCGCGCGCTGAGGGCGCATTCGCCCTGGA 1082
QY 1274 ggaatggtcaccagatctctgacttctcatctgagcccggtgcgcagatcagcaaaa 1333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1083 AGCCCTTTCGACCGATTCGCCGACTTCTGCCCAACGACGCGCAAAATGCGCTACCA 1142
QY 1334 gacgcagctgcagcggtgcagcagcactccctct 1370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1143 CCGCAGCGACTGTTGGCGCTCAAGAACCTGCGCTG 1179

RESULT 10
ID T58553
AC T58553 standard; cDNA; 4496 BP.
DF 02-APR-1997 (first entry)
DE Streptomyces pristinaespiralis snbA and papa intergenic region.
KW Streptogramin B; antibiotic; biosynthesis; pristinamycin;
KM virginiamycin; piperolic acid; cyclodeamination; papa; snbA; snbF;
PI p1pA; 3-hydroxypicolinic acid; hydroxylation; ds.
OS Streptomyces pristinaespiralis.
FH Key Location/Qualifiers
FT cds complement (1..332)
FT /*tag= a
FT /product= snbA
FT /note= "N-terminal coding region only, i.e. a
FT partial open reading frame"
FT 607..1674
FT cds
FT /*tag= b
FT /product= p1pA
FT 1800..2996
FT /*tag= c
FT /product= snbF
FT 3018..4496
FT /*tag= d
FT /product= Papa
FT /note= "N-terminal coding region only, i.e. a
FT partial open reading frame"
```

```
MO9601901-A1.
PD 25-JAN-1996.
PE 04-JUL-1995; F00889.
PR 08-JUL-1994; FR-008478.
```

PA (RHON ) RHONE POULENC RORER SA.  
PI Barriere JC, Blanc V, Blanche F, Crouzet J;  
PI Debussche L, Paris JM, Thibaut D, Bamas-Jacques N;  
PI Dutric-Rosset G, Famechon A;  
DR WPI: 96-097631/10.  
DR P-PSDB: W11584, W11585.  
PT New streptogramin B derivs. useful as antibiotics - produced by new  
PT mutants of Streptomyces having altered genes for streptogramin B  
PT biosynthesis  
PS Example 1: Page 107-111: 146pp: French.  
CC The papa gene of S.pristinaespiralis is involved in the biosynthesis  
CC of 4-dimethylamino-L-phenylalanine (DMAPA), a precursor for  
CC pristinamycin 1A. Upstream of the papa gene, on the complementary  
CC strand, is the snbA gene coding for 3-hydroxypicolinic acid-AMP ligase.  
CC The region between these two genes was sequenced and two open reading  
CC frames were identified. The first (pipa) decodes to an amino acid  
CC sequence with homology to ornithine cyclodeaminase from Agrobacterium  
CC tumefaciens. The pipa gene product is likely to catalyse the cyclo-  
CC deamination of lysine, leading to production of pipicollic acid.  
CC Mutations in the pipa gene were shown to affect pipicollic acid  
CC synthesis but not the synthesis of 3-hydroxypicolinic acid. The second  
CC open reading frame (snbA) could be decoded to give a product with  
CC homology to hydroxylases of the cytochrome P450 type.  
CC Disruption of the pipa and snbA genes can be used to produce  
CC strains of S.pristinaespiralis which are unable to produce the  
CC antibiotic pristinamycin I but which may be able to produce new,  
CC modified forms of it.  
SQ Sequence 4496 BP; 659 A; 1980 C; 1321 G; 536 T;

Query Match 5.9%; Score 83.4; DB 1; Length 4496;  
Best Local Similarity 45.7%; Pred. No. 8.8e-13;  
Matches 291; Conservative 0; Mismatches 346; Indels 0; Gaps 0;

OY 734 ctgcgacgagcgaagagcgctctacgactatctgataccgatactgagcaagcag 793  
DB 2342 CGTCAAGAACCCATGCGCGCATGACGCTCACTCCACGACACGACGACGCGCG 2401  
OY 794 gcgaagcgcggaacccagcgtatcagcgtctgtccgaagcgagtcgaatggcgacc 853  
DB 2402 CGCGCGCGCGGACGACGACGCTGATCTCCGACATCTGCGCGCGCGCGCGCGGACG 2461  
OY 854 gataccagctgacgaagcgaagagatgctgacgctgacgctgagcgctgagatc 913  
DB 2462 ACTACCGACGACGACGATCTCCGAATTCGGGCGCTGCTGATGACGCGCACGCTCTC 2521  
OY 914 ggtgtcgaattctcagctcagctgagctgctcgtgcgaagaagcccgagatcgcca 973  
DB 2522 CACTCTCATCTGCTGCGGCAACCGTGTGCTGTGAAGACCCCGGCGCGAAGGC 2581  
OY 974 ggaagctgacgaagcgtcccgagcgtatccagcgcgtctgaggaatactaccgagcctt 1033  
DB 2582 CGCGCGCGCGGACGACGCGCTGATCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2641  
OY 1034 ctgcgtggttgcgagatgagcgcatcctacacctcagattcgaattcattgagcgatgcaact 1093  
DB 2642 GCGCGCGGATCACCGTCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2701  
OY 1094 gaagaaagtgagcagacatctgtacgcgagatgctgctgctcgatgtagcggaataa 1153  
DB 2702 CATCTCCCGCGGAGCGATGCTGCGCTCTCCCTCTCCGCAACACGACGAAACGAGT 2761  
OY 1154 cgcgctgcgagatgacgctgacttcaagtcgcaaaagtttccacacaccccttggcca 1213  
DB 2762 CTTCACCGAGCCCGACACCTCGACCTCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCA 2821  
OY 1214 cggcagacatctgtgcttggcagaacactgcccgcggaataatcgtlcaacctctaa 1273  
DB 2822 CGGCATTCACACTACTGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2881  
OY 1274 ggaatggtcgcgacgagatctctgacttctcattgcccgcggtggtccagattcagcaaa 1333  
DB 2882 AGCCCTCTTGACGCGATTCGCCGACTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCA 2941

OY 1334 gaggcgatcgtcagcgagcgtgacgacatccctctg 1370  
DB 2942 CCGCGACGCGACTCTTGCGCTCAAGAACTGCGCGCTG 2978

## RESULT 11

ID V21187 standard; DNA; 53789 BP.

AC V21187;

DT 24-JUL-1998 (first entry)

DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.

KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;

OS Amycolatopsis mediterranei.

FH Key Location/Qualifiers

FT CDS 1825..15543

FT /tag= a

FT /label= ORF\_A

FT /product= "polyketide synthase"

FT CDS 15550..30759

FT /tag= b

FT /label= ORF\_B

FT /product= "polyketide synthase"

FT CDS 30895..36060

FT /tag= c

FT /label= ORF\_C

FT /product= "polyketide synthase"

FT CDS 36259..41325

FT /tag= d

FT /label= ORF\_D

FT /product= "polyketide synthase"

FT CDS 41373..51614

FT /tag= e

FT /label= ORF\_E

FT /product= "polyketide synthase"

FT CDS 51713..5293

FT /tag= f

FT /label= ORF\_F

FT /product= "polyketide synthase"

PN MO98078668-A1.

PD 26-FEB-1998.

PF 18-AUG-1997; E04495.

PR 20-AUG-1996; EP-810551.

PA (NOVS ) NOVARTIS AG.

PI Engel N, Schupp T, Toupet C;

PT WPI: 98-169172/15

DR P-PSDB: W52845-W52850.

PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used

PT to produce rifamycin and rifamycin analogues

PS Claim 4; Page 53-102; 205pp: English.

CC The present sequence represents a Amycolatopsis mediterranei rifamycin

CC synthesis gene cluster DNA fragment from the present invention. The

CC DNA fragment comprises a DNA region involved directly or indirectly

CC in the gene cluster responsible for rifamycin synthesis, including

CC the adjacent DNA regions to the right and left which, by reason of

CC their function in connection with rifamycin biosynthesis, qualify

CC as constituents of this rifamycin gene cluster, and functional

CC fragments, derivatives or constituents of these. The Amycolatopsis

CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used

CC for producing rifamycin, rifamycin analogues or precursors. It can also

CC be used for inactivating or modifying genes involved in ansamycin or

CC rifamycin biosynthesis. The DNA can be used for constructing mutant

CC actinomycetes strains from which the natural rifamycin or ansamycin

CC biosynthesis gene cluster has been partly or completely deleted. The

CC DNA fragment can be used for assembling a library of polyketide

CC syntheses, which can be used for assembling a library of polyketides.

CC A hybridisation probe of the invention can be used for identifying DNA

CC fragments involved in the biosynthesis of ansamycins.

SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T;

Query Match 5.5%; Score 77.8; DB 1; Length 53789;





CC Cytochrome P450soy was purified from *S. griseus* ATCC 13273. Two  
 CC similar forms of P450soy were isolated. P450soy-delta, is derived  
 CC from P450soy by in vitro proteolysis during isolation. One of the  
 CC tryptic peptide fragments of cytochrome P450soy and of of the  
 CC P450soy-delta protein were subjected to automated degradation. The  
 CC NH2 terminal sequences are given in R8306 and R8307. A mixture of  
 CC oligos that consist of possible DNA sequences that could encode the  
 CC MAS FGWHDCL of the tryptic peptide was made. It consists of the  
 CC sequences in Q43290-23. The oligo mixture was end-labeled and used  
 CC to probe the EMBL4 library of *S. griseus* DNA. Hybridising plaques  
 CC were isolated and a 4.8kb SacI DNA fragment was isolated from one  
 CC clone that hybridised to the oligo probe mixture. As segment of the  
 CC 4.8kb fragment was sequenced and found to contain an ORF. Within  
 CC this ORF was a section that corresp. exactly to the AA sequence  
 CC determined from the cytochrome P450soy tryptic peptide (see Q45569,  
 CC R8309). The gene encoding the P450soy protein was called soyC. Five  
 CC nucleotides downstream from the stop codon for soyC another ORF was  
 CC identified. This ORF encodes an apparent ferredoxin-like protein.  
 CC The gene was designated soyB and the protein ferredoxin-soy.  
 CC Sequence 1700 BP; 254 A; 636 C; 574 G; 236 T;

Query Match 5.2%; Score 73.6; DB 1; Length 1700;  
 Best Local Similarity 47.5%; Pred. No. 2.5e-10;  
 Matches 251; Conservative 0; Mismatches 274; Indels 3; Gaps 1;

QY 743 ggcacaagagcgctcctacgactatctgatacagatcatcagacaagcgagaagcc 802  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 742 GGGCCGGAGAGACTCGANAGATNCTGGGGCGGCTATGACCGAAGAGGGGGAGCC 801  
 QY 803 gggaaacgaagctatcagacatcgtltgccaagcgcaagtgcaatggcgacgaatcacaag 862  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 802 GGGTGAAGGCGCTCTGAGACAGCTGATCCACCGGAGCACCGGAGCGGGGTGACCG 861  
 QY 863 tgcagaaagccaagaagatgctgtgctgtactgctgtgctgtgatacgtgtcaaa 922  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 862 CGAAGACGCTGTCCTTCCTCCCGTCATCTCTCATGCGCGGCGACAGACGAGCGGCAA 921  
 QY 923 ttctcctagctcagacatgtagtctcctgccaagaagcccgagacatgcaggaagctgat 982  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 922 CATGATTCGCTCGGCGACGTTCAAGCTGCTGAGCACCACCGAAGCTGGCGGCTGCG 981  
 QY 983 cgaagctcccgagcgtatctcacaagcgcttgagagaactacatccgagctctcgtcgt 1042  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 982 GGGCGGGGAGACGAGCACCCTGGTGTGAGAGAGCTCTGCGGTCTCTCCATCGC 1041  
 QY 1043 tgcgcatgagcgcgacatccatccctcgatc---gaattcatggtgtgcaactgaagaa 1099  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1042 CGAAGGCGCTTCAGCGCTTGCGCAGCAGATGAGTGCAGCGCGCGACGATCCGCAA 1101  
 QY 1100 aggtgacgaatcctcgtctacgcagagatgctgtcgtcgtgtagagcgaaagcgtcgt 1159  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1102 GGGGGAGGGGCTGCTTCTTGACCTCGCTGATCAACCGGACGCGGACGCTGTTCCCGC 1161  
 QY 1160 ccgagatgcagctgactcagctcagcaaaagttcacacacacacacacacacacacacac 1219  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1162 GGGCGAGACACTGACGTGGAGGAGCGCCCGCCGCGCATACACTCGCTGCGCTTGGAGT 1221  
 QY 1220 ccatctgtcgtcgtgccaagcactgagcgcgcgaggaatcaltcgtcaac 1267  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1222 CCACAGTGGCTGGCGCAGAACTGGCGCGCGCGAGACTGACATCGC 1269

RESULT 15  
 Q14548  
 ID 014548 standard; DNA; 1400 BP.  
 AC 014548;  
 DT 28-JAN-1992 (first entry)  
 DE EryF gene.  
 KM C-6 hydroxylation: erythromycin; 6-deoxyerythromycin; antibiotics;  
 OS saccharopolyspora; cytochrome P450 monooxygenase; ss.  
 FH Saccharopolyspora erythraea.  
 key Location/Qualifiers

FT cds 158..1372  
 FT /\*tag= a  
 PT WO9116334-A.  
 PD 31-OCT-1991.  
 PF 16-APR-1991; 002600.  
 PR 18-APR-1990; US-510483.  
 PA (ABBO ) ABBOTT LABORATORIES.  
 PI Weber JM.  
 DR WPI: 91-339744/46.  
 PT P-PSDB; R14724.  
 PT New 6-deoxyerythromycin derivs. - are antibiotics with increased  
 PT acid stability, produced by cultivation of saccharopolyspora.  
 PS Disclosure; Fig 3; 56pp; English.  
 CC The eryf gene encodes the 6-hydroxylase component of the cytochrome  
 CC P450 monooxygenase system responsible for the hydroxylation of 6-de-  
 CC oxyerythronolide B to erythronolide B. Interruption of this step  
 CC results in the formation of deoxyerythromycin A and new derivatives  
 CC useful as antibiotic which have better stability against acids than  
 CC the corresponding erythromycins. Interruption of the reaction can  
 CC be effected by an insertion into the eryf gene of a plasmid, gene  
 CC replacement or chemical or light-induced mutagenesis.  
 CC The gene is positioned between the eryh and eryg genes.  
 CC See also Q14549.  
 SQ Sequence 1400 BP; 212 A; 495 C; 479 G; 214 T;

Query Match 4.9%; Score 68.6; DB 1; Length 1400;  
 Best Local Similarity 44.6%; Pred. No. 5e-09;  
 Matches 407; Conservative 0; Mismatches 494; Indels 12; Gaps 3;

QY 421 tccctcgtaagccgagcaagcctacgactcattccacactcgtatgatacgcgcgagc 480  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 390 TCCCCGAGAGACTGCGGAGACTACTCTGCCACCAATGGGCACGACGCCGCCGCC 449  
 QY 481 agcgcaagttcgtgcgtcgtgccaaccaagtggtgcaatgcggtgtgataagcttg 540  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 450 ACACCCGCGTCCGCAACCTGCTGTCGACAGATTACACCGTCCCGCGGTGAGGCGATGC 509  
 QY 541 agaacggaatcccaagagctgctcgtcgtcgtatcgagaacctgcgcgcgcaagaagat 600  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 510 GGGCCCGGTGAGCAATACCGCGGAGCTGCTCGACGAGGTGGAGTCCGCGCTGG 569  
 QY 601 gcaactcagcaggaactacgcaagcctcccgatacgatctatgctcgtcag 660  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 570 TCGACATCGCTTACCGCTTGTGGCCACCGCTGCCATCAAGATGATGTGGAGCTGCTCG 629  
 QY 661 gctaccggaagaagatalcccgactl-----gaaatacctaagcagatgataccc 714  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 630 GCGTGCAGAGAAATACCGCGGGGAGTTCGGCGGTGAGACTCGGAGATCCTGGTATGG 689  
 QY 715 gtccgcatgcaagatgaccttcgcaagagcgaagagcgctctacgactatcgtatc 774  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 690 ACCCGGAGCGGCGGAACACCGCGGCGGCGGCGCAGGAGGTGTCATCTTATCTTCG 749  
 QY 775 cgtatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 834  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 750 ACCTGTGAGCGCGCGCGACGAGCGCGCGCGACGACTGCTGCTCGCCTTATTCAGG 809  
 QY 835 gccag---gtcaaltggtgcagcagatcacagtgacgaagccaagaagatgtgtgacctgt 891  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 810 TTCAGAGCAGCAGATGACGTGCGGTACGCCCGCAGCGAGCTGTCTCCATCGCGCTGTGC 869  
 QY 892 tactgttcggtcggtcgtgataagtgtaattcctcagctcagctcagatgagatgagttcgtg 951  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 870 TGTGTGCGCGGCTTTCGAGGCGTGTGAGCTCTCATCGGATCGGACCTACTCTGTGC 929  
 QY 952 ccaaaagccggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1011  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 930 TCACCCACCGGAGACACTGCGCTGTGCGGCGGAGACCGGTGCGGCTCCCAACGCCG 989  
 QY 1012 gcgagaactactcgcgctctc---gctgtgtcgcagatgagcagatcctcactcgcg 1068  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 990 TCGAGGAGATCTGCGCTATACATCGCTCTCGAGAGACACACCGCGCTTGGCGGGAGG 1049





Db 241 ttcttcagcgaatgcccgttcacccctcgttgaaagccggcgaaagccctacgacttacc 300  
Qy 459 aacctgagatggtccgcccaggaacgagcgaatttcgtgcgtggcccaaccaagtgttggc 518  
Db 301 aacctgagatggtacccgcccaggaacgagcgaatttcgtgcgtggcccaaccaagtgttggc 360  
Qy 519 atgcccgtggtgataaacttgaaagacccggaatccagagagcgtgctgtctgtcag 578  
Db 361 atgcccgtggtgataaacttgaaagacccggaatccagagagcgtgctgtctgtcag 420  
Qy 579 agcctgcccgcgaagaaatgcaacttccacgaagactacgacgaacccctcccgata 638  
Db 421 agcctgcccgcgaagaaatgcaacttccacgaagactacgacgaacccctcccgata 480  
Qy 639 cgcactctcgtctgctcgaaggttaccggaagaagatatcccgacttgaataacta 698  
Db 481 cgcactctcgtctgctcgaaggttaccggaagaagatatcccgacttgaataacta 540  
Qy 699 acggaatcagatgacccggtccggaatgagcagatgacgttcgcagaaagcgaagcgctc 758  
Db 541 acggaatcagatgacccggtccggaatgagcagatgacgttcgcagaaagcgaagcgctc 600  
Qy 759 tacgactatctgataccgcatcatcgaaacgaagcgaagccgggaacgacgctatc 818  
Db 601 tacgactatctgataccgcatcatcgaaacgaagcgaagccgggaacgacgctatc 660  
Qy 819 agcatcgttgcgaacgagcaggttcaatgagccacgcatcacacgaagagcaagag 878  
Db 661 agcatcgttgcgaacgagcaggttcaatgagccacgcatcacacgaagagcaagag 720  
Qy 879 atgtgtgagcgttactgtgtgcgcccgttgatacgtgtgtcaatttctcagcttacc 938  
Db 721 atgtgtgagcgttactgtgtgcgcccgttgatacgtgtgtcaatttctcagcttacc 780  
Qy 939 atggaattccggtccgaagaagcccggaagatcgccgaagagcgtatcgaaagcgtccgaagcgt 998  
Db 781 atggaattccggtccgaagaagcccggaagatcgccgaagagcgtatcgaaagcgtccgaagcgt 840  
Qy 999 attcgaagcgtctgcgaagaacttaccggtcgtctcgtgtgtgcgagatgagccgcatc 1058  
Db 841 attcgaagcgtctgcgaagaacttaccggtcgtctcgtgtgtgcgagatgagccgcatc 900  
Qy 1059 ctccacgtcgaatcaggaattcattgtgcgttgcaactgaaagaaggtgacacgactcgtcta 1118  
Db 901 ctccacgtcgaatcaggaattcattgtgcgttgcaactgaaagaaggtgacacgactcgtcta 960  
Qy 1119 ccgcaagatgctgtctgtgcgttgatgagcgcgaaacgccttcccgatgcaagctgacttc 1178  
Db 961 ccgcaagatgctgtctgtgcgttgatgagcgcgaaacgccttcccgatgcaagctgacttc 1020  
Qy 1179 agtcgcaaaaggtttcacacacacaccccttggccacggaagcgaatcgttgccttggcag 1238  
Db 1021 agtcgcaaaaggtttcacacacacaccccttggccacggaagcgaatcgttgccttggcag 1080  
Qy 1239 caactgcccgcgaggaatcatcgttccacacgaagaaatggtgtgacgaagatcttctgac 1298  
Db 1081 caactgcccgcgaggaatcatcgttccacacgaagaaatggtgtgacgaagatcttctgac 1140  
Qy 1299 ttcttcacatgcccgggtgcccagaatcagcaaaagcgaatcgttgcggtgag 1358  
Db 1141 ttcttcacatgcccgggtgcccagaatcagcaaaagcgaatcgttgcggtgag 1200  
Qy 1359 gcactccctcgtctgtgagtcggcggaactccaaagcgta 1400  
Db 1201 gcactccctcgtctgtgagtcggcggaactccaaagcgta 1242

RESULT 2  
US-09-068-132-2  
; Sequence 2, Application US/09068132  
; Patent No. 611761  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Luet-Lok

; APPLICANT: Filesch, Sabine Lahja  
; APPLICANT: Nickerson, Darren Paul  
; APPLICANT: Hart, Alwyn James  
; TITLE OF INVENTION: Mutant Mono-Oxygenase Cytochrome P450C<sub>AM</sub>  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/09/068.132  
; CURRENT FILING DATE: 1998-09-14  
; EARLIER APPLICATION NUMBER: GB 95 22407.7  
; EARLIER FILING DATE: 1995-11-01  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Microsoft Word Office 98 for Macintosh  
; SEQ ID NO 2  
; LENGTH: 1242  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; NAME/KEY: P450cam C334A  
; LOCATION: 1 1242  
; OTHER INFORMATION: Mutant with Cys-334 (Tgc at nucleotides 1000 1002)  
; OTHER INFORMATION: substituted with Ala  
US-09-068-132-2

Query Match 88.4%; Score 1238.8; DB 5; Length 1242;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1240; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 159 agcgtgaaacacatatacaagaacgacccatcttgcctctgcacacccatgtgcagag 218  
Db 1 agcgtgaaacacatatacaagaacgacccatcttgcctctgcacacccatgtgcagag 60  
Qy 219 caactggtatcgcgacttcgacatgtacatccgttgaatcgtctgcgtgggtcagag 278  
Db 61 caactggtatcgcgacttcgacatgtacatccgttgaatcgtctgcgtgggtcagag 120  
Qy 279 gcttggagcgttctgcaagaatcaaacgttaccggaatcgtgtgtgactcgttgcgaagcgc 338  
Db 121 gcttggagcgttctgcaagaatcaaacgttaccggaatcgtgtgtgactcgttgcgaagcgc 180  
Qy 339 ggaactgtgacgcgacccgcgcgcgacacgtatccgttgagagcctatgaagataccgcac 398  
Db 181 ggaactgtgacgcgacccgcgcgcgacacgtatccgttgagagcctatgaagataccgcac 240  
Qy 399 ttcttcagcgaatgccccttccatccctcgttgaaagccggcgaaagcctaagacttacc 458  
Db 241 ttcttcagcgaatgccccttccatccctcgttgaaagccggcgaaagcctaagacttacc 300  
Qy 459 aacctgagatggtccgcccaggaacgagcgaatttcgtgcgtggcccaaccaagtgttggc 518  
Db 301 aacctgagatggtccgcccaggaacgagcgaatttcgtgcgtggcccaaccaagtgttggc 360  
Qy 519 atgcccgtggtgataaacttgaaagacccggaatccagagagcgtgctgtctgtcag 578  
Db 361 atgcccgtggtgataaacttgaaagacccggaatccagagagcgtgctgtctgtcag 420  
Qy 579 agcctgcccgcgaagaaatgcaacttccacgaagactacgacgaacccctcccgata 638  
Db 421 agcctgcccgcgaagaaatgcaacttccacgaagactacgacgaacccctcccgata 480  
Qy 639 cgcactctcgtctgctcgaaggttaccggaagaagatatcccgacttgaataacta 698  
Db 481 cgcactctcgtctgctcgaaggttaccggaagaagatatcccgacttgaataacta 540  
Qy 699 acggaatcagatgacccggtccggaatgagcagatgacgttcgcagaaagcgaagcgctc 758  
Db 541 acggaatcagatgacccggtccggaatgagcagatgacgttcgcagaaagcgaagcgctc 600  
Qy 759 tacgactatctgataccgcatcatcgaaacgaagcgaagccgggaacgacgctatc 818  
Db 601 tacgactatctgataccgcatcatcgaaacgaagcgaagccgggaacgacgctatc 660  
Qy 819 agcatcgttgcgaacgagcaggttcaatgagccacgcatcagatgaaagcgaagag 878



```
Db 661 agcaltgttgcacaagccaggltcaatggtgcagccagctaccagatgacgaagccaagaag 720
Qy 879 atgtgtgacctgtactgtgtcgccgacctgtgatacgtgtgtacttccctcaagcttcaac 938
Db 721 atgtgtgacctgtactgtgtcgccgacctgtgatacgtgtgtacttccctcaagcttcaac 780
Qy 939 atggaagttccttgccaaaagcccgagacatcgcacagagctgtatcgagcgtlcccgagcgt 998
Db 781 atggaagttccttgccaaaagcccgagacatcgcacagagctgtatcgagcgtlcccgagcgt 840
Qy 999 attcagacgcgtctgtcgagaagaactactcgcgcgtctctcgtctgttcgcgatagtgcgcgac 1058
Db 841 attcagacgcgtctgtcgagaagaactactcgcgcgtctctcgtctgttcgcgatagtgcgcgac 900
Qy 1059 ctccacctcgcgtatgaatttcatatgacctgtgcacactgaaagagtgagacagatctctgtcta 1118
Db 901 ctccacctcgcgtatgaatttcatatgacctgtgcacactgaaagagtgagacagatctctgtcta 960
Qy 1119 ccgcagatgtctgtcgtgcgtgtgatgacgcgcgaaaacgcctgcgcagatgcacatcgcacttc 1178
Db 961 ccgcagatgtctgtcgtgcgtgtgatgacgcgcgaaaacgcctgcgcagatgcacatcgcacttc 1020
Qy 1179 agtcgcacaagaagttcacacaacacacttgcgcacagcgagcactctgttcgtgcgacag 1238
Db 1021 agtcgcacaagaagttcacacaacacacttgcgcacagcgagcactctgttcgtgcgacag 1080
Qy 1239 caccgtgcgcgcgcgcgcgaatactatctgcacctcaaggaatgtgttaccagagttctctgac 1298
Db 1081 caccgtgcgcgcgcgcgcgaatactatctgcacctcaaggaatgtgttaccagagttctctgac 1140
Qy 1299 ttctccatgtcccgccgggtgcgccagatltcagcacagaagcgagatcgtcagcgcgctgcag 1358
Db 1141 ttctccatgtcccgccgggtgcgccagatltcagcacagaagcgagatcgtcagcgcgctgcag 1200
Qy 1359 gcactccctcgtctgtgagatccgcgcgactaccacaagcgcta 1400
Db 1201 gcactccctcgtctgtgagatccgcgcgactaccacaagcgcta 1242
```

```
RESULT 3
5212296-16
: Patent No. 5212296
: APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
: J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.
: TEPPERMAN, JAMES M.
: TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
: CYTOCHROMES
: NUMBER OF SEQUENCES: 19
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/569,781
: FILING DATE: 23-AUG-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 464,499
: FILING DATE: 12-JAN-1990
: APPLICATION NUMBER: 405,605
: FILING DATE: 11-SEP-1989
: SEQ ID NO:16:
: LENGTH: 1221
5212296-16
```

```
Query Match 6.3%; Score 88.2; DB 7; Length 1221:
Best Local Similarity 46.0%; Pred. No. 3.3e-14:
Matches 377; Conservative 0; Mismatches 433; Indels 9; Gaps 2;
```

```
Qy 466 tggatccgcccgcagagcgcagattcgtgcgttcgccaaccaagtggtgcatgacgcgg 525
Db 293 tggatccgcccgcagagcgcagattcgtgcgttcgccaaccaagtggtgcatgacgcgg 352
Qy 526 tgggtgataagctgagaaacgcgataccagagctgg---cctgtcgcgtgtatcgagagcc 582
Db 353 ggtatcaaggcattgcgcccgcagagtgagagtggtgcaacggtctcctcgcagcagatgac 412
```

```
Qy 583 tgcgcgcgcgaagagacagltgcacattcacccagagactacgcccgaaccttcccagatacgca 642
Db 413 tgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 472
Qy 643 tcttcaatgtctgtcgcgaagcttcaaccggaagaatatacccgcaattgaaataccctaagc 702
Db 473 tgaatgcgcgaactctcgtcgcgttcgccttaagccgacaacagatgtcttccagacgcgagaca 532
Qy 703 atcaatgaaccgcttcgcgatagtgcagatgaccttccagagagccaagagagcgctctacg 762
Db 533 agcggtgtgtgtcagltccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 592
Qy 763 actatcatatagacatcatcgcagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 822
Db 593 gttactctgagacgcgtctcatcaccaggttccaaaccggaacgcgcgcgcgcgcgcgcgcgcgcgc 652
Qy 823 tctgtgcacaagcgccagatgcaatggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 882
Db 653 ctctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 712
Qy 883 gtgcgcgtgttactgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 942
Db 713 cgaatgtgtcttctcatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 772
Qy 943 agttccttgccaaaagcccgagacatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1002
Db 773 taacctctgtgatacacaaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 832
Qy 1003 cagcgcctgtgcgaagaactactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1056
Db 833 ccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 892
Qy 1057 tcttcaacctcgcgatacagatgttcatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1116
Db 893 tgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 952
Qy 1117 taaccgagatgtctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1176
Db 953 tctgaacctcgcgtatgacaaaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1012
Qy 1177 tcaatgcacaagaagttcacacaacccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1236
Db 1013 tccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1072
Qy 1237 agcactgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1275
Db 1073 agaactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1111
```

```
RESULT 4
5212296-5
: Patent No. 5212296
: APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
: J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.
: TEPPERMAN, JAMES M.
: TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
: CYTOCHROMES
: NUMBER OF SEQUENCES: 19
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/569,781
: FILING DATE: 23-AUG-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 464,499
: FILING DATE: 12-JAN-1990
: APPLICATION NUMBER: 405,605
: FILING DATE: 11-SEP-1989
: SEQ ID NO:5:
: LENGTH: 1879
5212296-5
```

```
Query Match 6.3%; Score 88.2; DB 7; Length 1879:
Best Local Similarity 46.0%; Pred. No. 3.8e-14:
```





```
;
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: CR-9000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-7342
; TELEFAX: 302-892-7949
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-102-863-10

Query Match      5.2%; Score 73.6; DB 1; Length 1735;
Best Local Similarity 47.5%; Pred. No. 2.4e-10;
Matches 251; Conservative 0; Mismatches 274; Indels 3; Gaps 1;

QY 743 ggcgaagagagcgtctctacgtactatctatccgatacgcgaacgcgaagc 802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 ggcggcgacgacactcgagagatctggtggcgccgtgattcgaccgagagcgagcc 836

QY 803 gggaaacgcagctatcagcatcgltgcgaacggccaggtcaatggcgacgataccag 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 837 gggtgacggcctcctgagcagagctgattccaccgggacacccggagcgagcggtgacgc 896

QY 863 tgaagaaagcgaagagatgltggtcgttctgtgtggtggcggtcgatgagtgatcaa 922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 897 cgaaacagctggtcgttcgcccgtatccctgctatccgcccggacgagacgacggcgaa 956

QY 923 ttctctcagcttcaagatgagttctcagcgaagcccgagagacatcgcaagagatgat 982
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 957 catgattcgcgtcgcgacgattcagcgtctgtagccaccccgaaacgctggcgccctgcg 1016

QY 983 cgaagctcccgagcgatltccagccgcttgcgaggaactactcggcgcttcgtcgctgt 1042
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1017 ggcggcgcgagagacacacgcccgtggtgagagagctgctgcttctctccatcgc 1076

QY 1043 tgcgagtgccgcgcatcctcaactccgattac--gagtttcatgctggtgcaactgaa 1099
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1077 cgagggccttccagcgccctggcgacacgagacatgagagctgcggggcgagcattccgaa 1136

QY 1100 aggtgacacagatcctcgtaccgacagatgctgtcgtcgtgagatgagcgagaaagcgtg 1159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1137 gggagagggcggtgcttctgacctgctgatacmaacggcagacggcgagctgtccccc 1196

QY 1160 ccgagatgcagctgcagctcagctcgcgaagagtttcaacacacacaccccttggccagcgag 1219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1197 ggcggagacactcgactggagacggccggccgacatcagctcgttgcggttcggaggt 1256

QY 1220 ccaatctgacctggcgacgacgtgcgcgcggggaatcatcgtctac 1267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1257 ccaccagtgcttgggcgacgaactggccggcgagctgagacatcgc 1304

RESULT 8
PCT-US92-10885-10
; Sequence 10, Application PC/TUS9210885
; GENERAL INFORMATION:
; APPLICANT: SARIASLANI, SIMA
; TITLE OF INVENTION: CONSTITUTIVE
; TITLE OF INVENTION: EXPRESSION OF P450SOY
; TITLE OF INVENTION: AND FERREDOXIN-SOY IN
; TITLE OF INVENTION: STREPTOMYCES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
```

```
;
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 Inch,
; MEDIUM TYPE: 1.0 MB
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10885
; FILING DATE: 19921216
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: GALLEGOS, R. THOMAS
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: CR-9000-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-7342
; TELEFAX: 302-892-7949
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1735 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US92-10885-10

Query Match      5.2%; Score 73.6; DB 6; Length 1735;
Best Local Similarity 47.5%; Pred. No. 2.4e-10;
Matches 251; Conservative 0; Mismatches 274; Indels 3; Gaps 1;

QY 743 ggcgaagagagcgtctctacgtactatctatccgatacgcgaacgcgaagc 802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 ggcggcgacgacactcgagagatctggtggcgccgtgattcgaccgagagcgagcc 836

QY 803 gggaaacgcagctatcagcatcgltgcgaacggccaggtcaatggcgacgataccag 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 837 gggtgacggcctcctgagcagagctgattccaccgggacacccggagcgagcggtgacgc 896

QY 863 tgaagaaagcgaagagatgltggtcgttctgtgtggtggcggtcgatgagtgatcaa 922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 897 cgaaacagctggtcgttcgcccgtatccctgctatccgcccggacgagacgagcgagaa 956

QY 923 ttctctcagcttcaagatgagttctcagcgaagcccgagagacatcgcaagagatgat 982
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 957 catgattcgcgtcgcgacgattcagcgtctgtagccaccccgaaacgctggcgccctgcg 1016

QY 983 cgaagctcccgagcgatltccagccgcttgcgaggaactactcggcgcttcgtcgctgt 1042
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1017 ggcggcgcgagagacacacgcccgtggtgagagagctgctgcttctctccatcgc 1076

QY 1043 tgcgagtgccgcgcatcctcaactccgattac--gagtttcatgctggtgcaactgaa 1099
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1077 cgagggccttccagcgccctggcgacacgagacatgagagctgcagcgggcgagcattccgaa 1136

QY 1100 aggtgacacagatcctcgtaccgacagatgctgtcgtcgtgagatgagcgagaaagcgtg 1159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1137 gggagagggcggtgcttctgacctgctgatacmaacggcagacggcgagctgtccccc 1196

QY 1160 ccgagatgcagctgcagctcagctcgcgaagagtttcaacacacacaccccttggccagcgag 1219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1197 ggcggagacactcgactggagacggccggccgacatcagctcgttgcggttcggaggt 1256

QY 1220 ccaatctgacctggcgacgacgtgcgcgcggggaatcatcgtctac 1267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1257 ccaccagtgcttgggcgacgaactggccggcgagctgagacatcgc 1304

RESULT 9
5212296-17
; Patent No. 5212296
; APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
```





STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/444,623  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/095,734  
FILING DATE: 22-JUL-1993  
APPLICATION NUMBER: 07/711,334  
FILING DATE: 06-JUN-1991  
APPLICATION NUMBER: 07/367,894  
FILING DATE: 19-JUN-1989  
APPLICATION NUMBER: PCT/US90/03451  
FILING DATE: 18-JUN-1990  
APPLICATION NUMBER: PCT/US89/02962  
FILING DATE: 07-JUL-1989  
APPLICATION NUMBER: 07/361,944  
FILING DATE: 05-JUN-1989  
APPLICATION NUMBER: 07/223,089  
FILING DATE: 22-JUL-1988  
APPLICATION NUMBER: 07/216,390  
FILING DATE: 07-JUL-1988  
APPLICATION NUMBER: 07/163,546  
FILING DATE: 03-MAR-1988  
APPLICATION NUMBER: PCT/US88/00614  
FILING DATE: 29-FEB-1988  
APPLICATION NUMBER: 07/020,451  
FILING DATE: 02-MAR-1987  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WHI93-11M2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4394 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-444-623-1

Query Match 2.6%; Score 37; DB 3; Length 4394;  
Best Local Similarity 50.3%; Pred. No. 1.2;  
Matches 91; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 1220 ccatctgtgcttgcagccagccctgcccgcgggaatcatctgcacccctcaagaatg 1279  
DB 2066 CATTTGGGCTTCGGTCCGCCGCTCTAGAGCTCCGCCGCCGCCACCGCCGAGG 2125  
QY 1280 gctgacccagatctctgacttccatctgccccgggtgcccagatctcagaagaagcgg 1339  
DB 2126 GGTGTCTGTCTGGCGGCGACCTCCAAATCCCGAGGTCGCGGCGTCAAGAAATGCCGCGC 2185  
QY 1340 catctgcaagcggtgcagcgacacccctctggtctggatccggcgactcaacaagcgt 1399  
DB 2186 CGACGCGCGCAGCGTGGCGCCAGTGTGCTGTGACCAAGGTGGGGCGGCCCAACGAGCGGCG 2245  
QY 1400 A 1400  
DB 2246 A 2246

RESULT 14  
US-08-471-869-1  
Sequence 1, Application US/08471869  
Patent No. 6022745  
GENERAL INFORMATION:  
APPLICANT: Aldovini, Anna  
APPLICANT: Young, Richard A.  
TITLE OF INVENTION: Homologously Recombinant Slow Growing  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,869  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/08267  
FILING DATE: 22-JUL-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/095,734  
FILING DATE: 22-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/711,334  
FILING DATE: 06-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/367,894  
FILING DATE: 19-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US90/03451  
FILING DATE: 18-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US89/02962  
FILING DATE: 07-JUL-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/361,944  
FILING DATE: 05-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/223,089  
FILING DATE: 22-JUL-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/216,390  
FILING DATE: 07-JUL-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/163,546  
FILING DATE: 03-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US88/00614  
FILING DATE: 29-FEB-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/095,734  
FILING DATE: 22-JUL-1993  
APPLICATION NUMBER: 07/711,334  
FILING DATE: 06-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WHI93-11M2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 02:46:36 ; Search time 1168.17 Seconds  
(without alignments)  
5292.534 Million cell updates/sec

Title: US-09-246-451-1  
Perfect score: 1402  
Sequence: 1 ctgcagcagatcgtatccgcgt.....gcgcactaccaaagcgggtata 1402

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*

44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: em\_est20:\*  
50: em\_est21:\*  
51: em\_est22:\*  
52: gb\_est30:\*  
53: gb\_est31:\*  
54: gb\_est32:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: em\_est35:\*  
83: em\_est36:\*  
84: em\_est37:\*  
85: gb\_est48:\*  
86: gb\_est49:\*  
87: gb\_est50:\*  
88: gb\_est51:\*  
89: gb\_est52:\*  
90: gb\_est53:\*  
91: gb\_est54:\*  
92: gb\_est55:\*  
93: gb\_gss1:\*  
94: gb\_gss2:\*  
95: gb\_gss3:\*  
96: gb\_gss4:\*  
97: em\_gss1:\*  
98: em\_gss2:\*  
99: em\_gss3:\*  
100: em\_gss4:\*  
101: gb\_gss5:\*  
102: gb\_gss6:\*  
103: em\_gss5:\*  
104: em\_gss6:\*  
105: gb\_gss7:\*  
106: gb\_gss8:\*  
107: gb\_gss9:\*  
108: em\_gss7:\*  
109: em\_gss8:\*  
110: gb\_gss11:\*  
111: gb\_gss10:\*  
112: em\_gss9:\*  
113: em\_gss10:\*  
114: em\_gss11:\*  
115: em\_gss12:\*  
116: gb\_gss12:\*



KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 408)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Schellensberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The MashU-HMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of MedicineP 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:288584 Seq primer: -28M13 rev7 from Amersham High quality sequence stop: 373. Location/Qualifiers ..408 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:477840" /clone_lib="Scars mouse embryo NMMEI3.5 14.5" /sex="unknown" /tissue_type="embryo" /dev_stage="13.5-14.5dpc total fetus" /lab_host="DH10B" /note="Vector: pT773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACCAATCTGTAAGTGAGCGGGCCGCCGAATTATTTTATTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 J]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Felima Bernaldo."
BASE COUNT	76 a 127 c 123 g 82 t
ORIGIN	
Query Match	3.0%; Score 42.6; DB 20; Length 408;
Best Local Similarity	48.5%; Pred. NO. 0.51;
Matches 117; Conservative	0; Mismatches 124; Indels 0; Gaps 0;
OY	456 ccacaccgatgagatccgcgccgagcggcgcaatttcgtgctgagccaaccaagtgtgt 515 
Db	333 CCCACCTCTCTGGGGCCAGGACATTTTAGTCGCCAGCGGGGTGCCTTGCTGAAAGGCAC 274
OY	516 ggcacbgcgggtggtgataaactgtgaagaacaggatccagaagcttgacctgctgcgtatc 575 
Db	273 TACCTGCAGGTAGAAGTCTCCAGGCGGACGACGCTGCACAGGGAAGGGCCGCTAGCTGCAC 214
OY	576 gagagctgcgcgccgcaagagacagtgcacattcacccgaagactacgcgcgaaccttcccg 635 
Db	213 TACCACCTGCTCATGCACAAGCAGAGTGGCCAGCCCTCTGTGGAAGAAGAGAAATCCACTATA 154
OY	636 ataacacattcatgctgctgcaggtctaacggnaagaatatccgcgaacttgaaataac 695 
Db	153 CTGGGCAAGGCTTCTCTGCTGACCTTGGCAAGCAGGTGTTTAGCAGGATCACGAGAAATC 94
OY	696 c 696

	Db	93	C	93	
	RESULT#	3			
	LOCUS	M45899/c			
	DEFINITION	M45899	411 bp	mRNA	EST
	ACCESSION	mc78d12.r1	Soares mouse embryo	NBME13.5	14.5 Mus musculus cDNA
	VERSION	M45899			
	KEYWORDS	M45899.1	GI:1330606		
	SOURCE	EST.			
	ORGANISM	house mouse.			
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
	TITLE	1 (bases 1 to 411)			
	JOURNAL	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubugue,T.,			
	COMMENT	Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thelsing,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterson,R. The Mashu-HMI Mouse EST Project Unpublished (1996) On Apr 14, 1993 this sequence version replaced gi:837668. Contact: Marra M/Mouse EST Project Mashu-HMI Mouse EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information. MG1:226447 Seq primer: ETPrimr. Location/Qualifiers 1..411 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone IMAGE:354647" /clone_id="Soares mouse embryo NBME13.5 14.5" /sex="unknown" /diss_type="embryo" /dev_stage="13.5-14.5dpc total fetus" /lab_host="DH10B" /note="Vector: pRTT3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dN) primer 15' TGTTACCACTGTGAGTGGAGCGGCCGGAATTTTTTTTTTTTTTTTTTTT T 3'}, on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 ]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRTT3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."			
	BASE COUNT	78 a	127 c	124 g	82 t
	ORIGIN				
	Query Match	3.0%:	Score 42.6:	DB 91:	Length 411:
	Best Local Similarity	48.5%:	Pred. No. 0.51:		
	Matches 117:	Conservative	0:	Mismatches 124:	Indels 0:
	Gaps	0:			
	456	ccaccctgatgatgattcggccgcagagcgccagggttcgtgctgagccacaagaagtgtt	515		
	Db	332	cCCACTCTCTGGGGCCAGGCATTTCAGTCCAGCCGGGTGCTGGGCTGCTGAAAGGCAC	273	
	516	ggcatgccgttgtagtaagctgtagaagacagatccaggaagctgctgcctgcctgatac	575		

Db	212	TACGTGACGAGTAGAAGTCTCCACGAGGCGCAGCAGAGCTGCACGAGGAAGGCGCGTAGCTGCAC	213
Qy	576	gagagcctcgqcccgcaagagacatgtgcaacttcaccggaggaactagccgaaccctcccg	635
Dd	212	TACCACCTCGTCATTCACAGCAGAGAGTGGCCAGCCCTCGTGGAAGAAGACGAAATCCACTATA	153
Qy	636	atacgcatcttcacgtctgctgcagctacccgagaagaatatacccgacttgaaatac	695
Dd	152	CTGGGCAACAGGCTTCTCTCTCTGACCTTGGCAAGCAGGTGTTAGCAGAGGACACGAGAGC	93
Qy	696	c 696	
Dd	92	c 92	
RESULT	4		
CNS0175Y			
LOCUS	CNS0175Y		
DEFINITION	CNS0175Y 1101 bp DNA	GSS	26-JUL-1999
	Drosophila melanogaster genome survey sequence SP6 end of BAC		
	BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit		
	fly), genomic survey sequence.		
ACCESSION	AL108460		
VERSION	AL108460.1	GI:5628764	
KEYWORDS	GSS.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1101)		
AUTHORS	Genoscope.		
TITLE	Submitted		
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segreifgenoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector peloBAC11.		
COMMENT			
FEATURES			
source	location/Qualifiers		
	1..1101		
	/organism="Drosophila melanogaster"		
	/plasmid="peloBAC11"		
	/db_xref="taxon:7227"		
	/clone_lib="DrosBAC"		
	/clone="BACN37L08"		
	/note="end : SP6"		
BASE COUNT	254 a 176 c 160 g 152 t 359 others		
ORIGIN			

	Query Match	2.9%	Score 41	DB 123	Length 1101
	Best Local Similarity	13.2%	Pred. No. 1.7		
	Matches 51	Conservative 175	Mismatches 161	Indels 0	Gaps 0
OY	256	atctgtctccgcgcgtgcagagagccttggccagtlctgcaagaatcaacgctaccggatc	315		
Db	709	AKMACSSSSSSSSSSSSCASTRSSSSSRKGAWSSSCACSSGSGGAACAGCAGGGGGRGC	768		
OY	316	tgtgtgtgacctgcgtctgcacgagcgacgacatgcatgcgcaactgcgcgcaactgacctgt	375		
Db	769	GSGCCASGCVGCAACSSSSCASSMCGVSSGSGCSACSGCCGVSSCSAVSASASVSMS	828		
OY	376	aggcctatgaagattatcgcgcacatttccacgagatgcccgttatccctcgttgaaagcg	435		
Db	829	KVASAVSACSNAVSGMSAGAVSSSCRSSVASVSAASVSSSSSSSVSAASVMS	888		
OY	436	gcgaagcctacgactcattccacacctcgaatgagtcgcccgaagcgacgcagatttcgt	495		

Db 889 SSSASSSMVAALAAVAVSVSVASVSSSSSSSSSSSVVSAVSAAASVSSSS 948  
 Qy 496 cgcgtgcaccaaccaagtgtgtgtgcatagcgcggtgtgtgataagctggaagacgcgatccagg 555  
 Db 949 SSVSTSSASVSSVSASVSMASVAVSSSSSSSSSVSVSVAAASAAAAAASSSS 1008  
 Qy 556 agctgagctctgcgcgatcagagagctctgcgcgcgaagaagagcgaacttcaccagg 615  
 Db 1009 ASAAVAVVSSASSSSSSSSSSSSSSSVSVSSSVSVSVSSSVSVAAVASASASASV 1066  
 Qy 616 actagcgcgaaccttcaccataga 642  
 Db 1069 SSVAAAVAAAMAAAAASASASASVAA 1095

RESULT	5
LOCUS	AI188899
DEFINITION	AI188899 547 bp mRNA EST 28-JAN-1999
ACCESSION	GI19957.5prine GH Drosophila melanogaster head POT2 Drosophila melanogaster cDNA clone GH19957 5prime, mRNA sequence.
VERSION	AI188899
KEYWORDS	AI188899.1 GI:4202919
SOURCE	EST.
ORGANISM	fruit fly. Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 547)
AUTHORS	Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein, P., Lewis, S. and Rubin, G. M.
TITLE	BDBG/HMM Drosophila EST Project
JOURNAL	Unpublished (1997)
COMMENT	Contact: Harvey, D. G. M. Rubin-Molecular and Cell Biology University of California Berkeley 539 LSA, Berkeley, CA 94720-3200, USA Fax: 510 643 9947 Email: <a href="http://www.fruitfly.org/EST_estefruitfly.berkeley.edu">http://www.fruitfly.org/EST_estefruitfly.berkeley.edu</a> Plate: 199 row: E column: 9 High quality sequence stop: 540.
FEATURES	location/Qualifiers
source	1..547
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone="GH19957"
	/clone_lib="GH Drosophila melanogaster head POT2"
	/sex="male and female"
	/dev_stage="adult"
	/lab_host="DHS - alpha"
	/note="Organ: head; Vector: POT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."
BASE COUNT	135 a 153 c 171 g 88 t
ORIGIN	

Query Match	Similarity	Score	DB	Length
Best local	49.3%	49.38	Pred. No. 1.8	
Matches	106	Conservative	0	Mismatches 109; Indels 0; Gaps 0
QY	523	cagttgtagaataagcttggaaacccgataccagagaccttggcctctgcgtatcagagacc	582	
Db	260	CGAGAGTGAACACGAGGTGTGGCACCGCCATCGAATCATGTGGCCGGGAGATATCAGCCAGAGCC	319	
QY	583	tgcgcgcgaaggaacagtggcaacttcacgcgaagactacgcccgaacccctccgcataagca	642	
Db	320	GGCTGATGCTCGAGAGTATCCAGTCCGCTTCGATGGCTTCACACACGACCTCATCA	379	
QY	643	tctctatgcttggctgcaggtctacccgagagaagaatattccgcgaacttgaataacttaagc	702	
Db	380	CAATGATGGGTCTCGGCACCGCGGTGGAGGAGACCAAGAACATATATCGACATGAGC	439	











DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
Db	639	CGCCCTCAACACTGCGCAAGGTGGAAGCGCGTATGACACCTGCGCTGTGTTGGCGCTTGGGAGC	698							
Qy	905	ccctgatacagctgtgtcaattctcctaagcttcagcatcagcatgagcttcctgcccacaaagcccgga	964							
Db	699	TGTGGCGCTTGGCGCTGGGCTTGGCGCTGCGAAGAGCGCTGCGCTCAAGAATCTACGACATTG	758							
Qy	965	gcattgcacagagagcctgcatcgagc	987							
Db	759	ACATCAATCTCGACAAAGTTTGAGC	781							
RESULT 14	CNS004NB/C	CNS004NB	839 bp	DNA	GSS	03-JUN-1999				
		Drosophila melanogaster genome survey sequence TERN3 end of BAC #								
		BACR01E16 of RPCI-98 library from Drosophila melanogaster (fruit								
		fly), genomic survey sequence.								
		AL054280								
		AL054280.1	GI:4931788							
		GSS.								
		fruit fly.								
		Drosophila melanogaster								
		Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;								
		Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;								
		Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.								
		1 (bases 1 to 839)								
		Genoscope.								
		Direct Submission								
		Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;								
		BP 191 91006 EVRI cedex - FRANCE (E-mail : seqref@genoscope.cns.fr								
		- Web : www.genoscope.cns.fr)								
		Determination of this BAC-end sequence was carried out as part of a								
		collaboration with the Berkeley Drosophila Genome Project (BDGP).								
		The BDGP is constructing a physical map of the Drosophila								
		Melanogaster genome using these BACs. For further information								
		please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila								
		melanogaster BAC library was prepared by Kazutoyo Osoegawa and								
		Aaron Mammoser in Pieter de Jong's laboratory in the Department of								
		Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,								
		NY. The library is named RPCI-98 and was constructed by partial								
		EcoRI digestion of Drosophila DNA provided by the BDGP from the								
		isogenic strain y2; cn bw sp, the same strain used for the BDGP's								
		P1 and EST libraries. A more detailed description of the library								
		and how to order individual BAC clones, the entire library, or								
		filters for hybridization from the BACPAC Resource center can be								
		found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .								
FEATURES	Source	Location/Qualifiers								
		1..839								
		/organism="Drosophila melanogaster"								
		/db_xref="taxon:7227"								
		/clone_lib="RPCI-98"								
		/clone="BACR01E16"								
		/note="end : TERN3"								
BASE COUNT	285 a	67 c	77 g	27 t	383 others					
ORIGIN										
Query Match	2.6%	Score 37;	DB 122;	Length 839;						
Best Local Similarity	17.6%	Pred. No. 18;								
Matches	47;	Conservative 105;	Mismatches 115;	Indels 0;	Gaps 0;					

OY	438	gaagctcagctatccatccacccagtgatgcgcgcgagcagcgcaagttcgtcg	497
Db	644	SBTYSCCTTCKCKCKRYTCSYBTFTTTTSCBCBCSCSGSBTTBSCSSSSSSSGSTTTSSSSSS	585
OY	498	ctgycacacaaagtgtgtgcacgcg	524
Db	584	SGSKCSTSSSSSTSTSGSGSTSSCS	558
RESULT	15		
LOCUS	AV389986		
DEFINITION	AV389986 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii	EST	21-JAN-2000
ACCESSION	AV389986		
VERSION	AV389986.1	GI:5544202	
KEYWORDS	EST.		
SOURCE	Chlamydomonas reinhardtii.		
ORGANISM	Chlamydomonas reinhardtii		
REFERENCE	Chlamydomonas reinhardtii.		
AUTHORS	Eukaryota: Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;		
TITLE	Chlamydomonadaceae; Chlamydomonas.		
JOURNAL	1 (bases 1 to 411)		
COMMENT	Asamiu, E., Nakamura, Y., Sato, S., Fukuzawa, H. and Tabata, S.		
FEATRES	A large Scale Structural Analysis of cDNAs in a Unicellular Green		
SOURCE	Alga, Chlamydomonas reinhardtii. I. Generation of 3433		
	Non-redundant Expressed Sequence Tags		
	DNA Res. 6, 369-373 (1999)		
	On Jul 8, 1999 this sequence version replaced gi:5422438.		
	Contact: Yasukazu Nakamura		
	The First Laboratory for Plant Gene Research		
	Kazusa DNA Research Institute		
	Yana 1532-3, Kisarazu, Chiba 292-0812, Japan		
	Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.		
	Location/Qualifiers		
	1. 411		
	/organism="Chlamydomonas reinhardtii"		
	/strain="C9"		
	/db_xref="taxon:3055"		
	/clone="CM054f01.r"		
	/clone_lib="Chlamydomonas reinhardtii C9"		
	/dev_stage="photoautotrophic growth"		
	/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:		
	XhoI"		
BASE COUNT	69 a 153 c 128 g 61 t		
ORIGIN			
Query Match	2.6%; Score 36.8; DB 62; Length 411;		
Best Local Similarity	48.6%; Pred. NO. 18;		
Matches 101; Conservative 0; Mismatches 107; Indels 0; Gaps 0;			
OY	1046	cgaatgcccgcacatcctaccctccgaltacgaagttcaatggtcgatgcaactgaagaaagtga	1105
Db	197	CGCGGCGCTGACCCCTCTCCGAGAGCAAGCCTTACGCCAAGCTGGAAGAAAGAGACT	256
OY	1106	ccagttctgtctacgcgagatgtctcgtcgtgatgagcgcgaaacgctcgcgat	1165
Db	257	GAAAGCCCTGGAGAAAGCCCTGAAGCACTAGCAGGCGGACAGCCGCCCGCTGTGACCT	316
OY	1166	gcacgtcgtactcaatgcgcgaaggttcaacacacacaccttggccacgcagcaccat	1225
Db	317	GAAAGCCACACATGAGGCGGACCAAGCCCGCTTGCCCAACTACGCTAAGGCGGCGCTGCT	376
OY	1226	gtgccttggccagcaactcgtgcgcgcg	1253
Db	377	GTCGCGCAACGACGCTGTGCCCCACCTG	404
Search completed: October 4, 2000, 05:46:18			
Job time: 10782 sec			

Search completed: October 4, 2000, 05:46:18  
Job time: 10782 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 05:49:24 ; Search time 76.18 Seconds  
(without alignments)  
128,722 Million cell updates/sec

Title: US-09-246-451-2

Perfect score: 2180  
Sequence: 1 TTTETIOSNANLAPLPHPVPE.....IVSGVOALPLVMDPATTKAV 414

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2137	98.0	405	Y04128	Pseudomonas putida
2	1185	54.4	446	Y04126	Bacterial and mamm
3	365	16.7	396	R60777	Mycnamicin IV hyd
4	332	15.2	408	R77867	S. clavuligerus OR
5	330.5	15.2	587	W33274	S. fradiae tylosin
6	330	15.1	406	R11349	Cytochrome enzyme
7	329	15.1	411	W54389	Actinomadura hibis
8	324.5	14.9	398	R11585	Streptomyces prist
9	322	14.8	403	R11350	Cytochrome enzyme
10	316	14.5	412	R38309	Sequence of the p4
11	313.5	14.4	410	R51368	Protein containing
12	297.5	13.6	404	R14724	6-hydroxylase enco
13	190	8.7	398	R47521	Vitamin D hydroxyl
14	161	7.4	422	W36128	Daunomycin C-14 hy
15	161	7.4	474	W36132	N-terminal modifie
16	156	7.2	422	W00729	Daunorubicin 14-hy
17	145	6.7	518	W67616	A. nidulans phenyl
18	144	6.6	494	R62825	Human steroid-21-h
19	144	6.6	533	R15057	Cytochrome P450C25
20	129.5	5.9	1169	R76544	Mitochondrial cyto
21	129	5.9	512	W93216	Human cytochrome p
22	128	5.9	516	W67617	P. chrysogenum phe
23	127	5.8	508	W35711	Chrysanthum flavon
24	127	5.8	512	R72365	Human auxillary cy
25	127	5.8	512	R93172	Human cytochrome p
26	126	5.8	503	R72363	Human cytochrome p
27	126	5.8	503	R81464	Human derived cyto
28	126	5.8	503	R93170	Human cytochrome p
29	126	5.8	503	Y05202	Human CYP3A4 prote
30	126	5.8	512	R72366	Human auxillary cy
31	126	5.8	512	W00652	Cytochrome P450A1
32	126	5.8	512	R93173	Human cytochrome p
33	124.5	5.7	493	R72362	Human cytochrome p

34	124.5	5.7	493	1	R81467	Human derived cyto
35	124.5	5.7	493	1	R93169	Human cytochrome p
36	124	5.7	524	1	P70577	Rat hepato-cytochr
37	121.5	5.6	523	1	R59291	Rat liver cytochro
38	121.5	5.6	898	1	P61030	Entire coded seque
39	121.5	5.6	898	1	P61082	Complete translati
40	121.5	5.6	899	1	P61056	Translation of pla
41	121.5	5.6	1144	1	P81334	Expression prod. o
42	121.5	5.6	1150	1	P81335	Expression prod. o
43	121.5	5.6	1150	1	P81337	Expression prod. o
44	121.5	5.6	1162	1	P81336	Expression prod. o
45	121	5.6	512	1	R72364	Human auxillary cy

ALIGNMENTS

RESULT 1	
ID Y04128	Y04128 standard; Protein; 405 AA.
AC Y04128:	
DT 11-JUN-1999 (first entry)	
DE Pseudomonas putida cytochrome P450 protein P450-cam.	
KW Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;	
KM oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;	
OS bioremediation; environmental pollutant.	
OS Pseudomonas putida.	
PN M0908812-AL.	
PD 25-FEB-1999.	
PE 17-AUG-1998; U16979.	
PR 20-AUG-1997; US-056754.	
PA (UYRP ) UNIV ROCHESTER.	
PI Jones JP, Shimoi M;	
DI MPI, 99-190131/16.	
DR N-PSDB; X19926.	
PT New P450 fusion proteins - comprising a portion of a bacterial	
PT cytochrome P450 protein and a portion of a mammalian cytochrome P450	
PS protein	
PS Disclosure: Page 12-13: 51pp: English.	
CC The present invention describes a fusion proteins comprising a portion	
CC of a bacterial cytochrome P450 protein and also a portion of a mammalian	
CC cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or	
CC any compound having a carbon-hydrogen bond. The fusion protein can be	
CC used for hydroxylating a compound to be oxidised. It can also be used in	
CC the bioremediation of an environmental pollutant. Since the fusion	
CC protein is soluble, it can be subject to structural elucidation by X-ray	
CC crystallography for designing functional proteins. It can be readily	
CC expressed in soil bacteria to facilitate bioremediation. The present	
CC sequence represents Pseudomonas putida cytochrome P450 protein P450-cam	
CC from the present invention.	
SO Sequence 405 AA:	
Query Match 98.0%; Score 2137; DB 1; Length 405;	
Best Local Similarity 100.0%; Pred. No. 9, 2e-207;	
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY 10 NLAPLPHPVDEHLVFDEDMYNPSNLGSGVOEAWANVLQESNVPDLVWTRKNGHMIATRGQ 69	
DB 1 NLAPLPHPVDEHLVFDEDMYNPSNLGSGVOEAWANVLQESNVPDLVWTRKNGHMIATRGQ 60	
OY 70 LIREAEDVYHPSSECFPIREAGCEAVDTFTSGMDPPEQOFALANOVVGMVVDKLEN 129	
DB 61 LIREAEDVYHPSSECFPIREAGCEAVDTFTSGMDPPEQOFALANOVVGMVVDKLEN 120	
OY 130 RIEBLACSLIESLRPOGQCFNTEDEYAEPPRIPIFMLLAGIPEDIPILKXLTQOMTRPDG 189	
DB 121 RIEBLACSLIESLRPOGQCFNTEDEYAEPPRIPIFMLLAGIPEDIPILKXLTQOMTRPDG 180	
OY 190 SMTFAEKALYLYLPIIEORRQKPGTDAISIVANGVNGRPTSDSEAKRMCGLLVVG 249	
DB 181 SMTFAEKALYLYLPIIEORRQKPGTDAISIVANGVNGRPTSDSEAKRMCGLLVVG 240	

OY 250 LDTVNFSLFSMEFLAKSPENROELIERPERIPACEELLRRSLVADGRILTSDEYFHG 309  
 |||||  
 DB 241 LDTVNFSLFSMEFLAKSPENROELIERPERIPACEELLRRSLVADGRILTSDEYFHG 300  
 |||||  
 OY 310 VOLKKDQILLPQMLSGDLDERENACPMHVDPSQKVSHTTFGSHLCLGQHLARRRITV 369  
 |||||  
 DB 301 VOLKKDQILLPQMLSGDLDERENACPMHVDPSQKVSHTTFGSHLCLGQHLARRRITV 360  
 |||||  
 OY 370 TLKEMLTRIPDPSIAPCAQIOHKSGIVSGVQALPLVMDPATTKAV 414  
 |||||  
 DB 361 TLKEMLTRIPDPSIAPCAQIOHKSGIVSGVQALPLVMDPATTKAV 405  
 |||||

RESULT 2  
 Y04126  
 ID Y04126 standard; Protein: 446 AA.  
 AC Y04126:  
 DT 11-JUN-1999 (first entry)  
 DE Bacterial and mammalian chimeric cytochrome P450 protein.  
 KW Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;  
 KM oxidase; hydrocarbon; carbon-hydrogen bond; hydroxylating;  
 KW bioremediation; environmental pollutant.  
 OS Synthetic.  
 PN MO9908012-A1.  
 PD 25-FEB-1999.  
 PF 17-AUG-1998; 016979.  
 PR 20-AUG-1997; US-056754.  
 PA (UYRP ) UNIV ROCHESTER.  
 PI Jones JP, Shimoji M;  
 DR WPI: 99-190131/16.  
 DR N-PSDB; X19916.  
 PT New P450 fusion proteins - comprising a portion of a bacterial  
 cytochrome P450 protein and a portion of a mammalian cytochrome P450  
 protein.  
 PS Claim 24: Page 6-8: 51pp; English.  
 CC The present sequence is a fusion proteins comprising a portion of a  
 bacterial cytochrome P450 protein and also a portion of a mammalian  
 cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or  
 any compound having a carbon-hydrogen bond. The fusion protein can be  
 used for hydroxylating a compound to be oxidised. It can also be used in  
 the bioremediation of an environmental pollutant. Since the fusion  
 protein is soluble, it can be subject to structural elucidation by X-ray  
 crystallography for designing functional proteins. It can be readily  
 expressed in soil bacteria to facilitate bioremediation.  
 SQ Sequence 446 AA;

Query Match 54.4%; Score 1185; DB 1; Length 446;  
 Best Local Similarity 57.6%; Pred. No. 6, 6e-111;  
 Matches 253; Conservative 33; Mismatches 99; Indels 54; Gaps 9;

OY 10 NLAPRLPHVPEHLVFPDQMYNPSNLGAGVOEAMVLOESNVPLVMTKRCGHWMTATRCQ 69  
 |||||  
 DB 1 NLAPRLPHVPEHLVFPDQMYNPSNLGAGVOEAMVLOESNVPLVMTKRCGHWMTATRCQ 60  
 |||||  
 OY 70 LIREAVEDYKHFSSECPFIPREAGEAYDIFPTSMDDPEQROFRALANOVVGMFVVDKLEN 129  
 |||||  
 DB 61 LIREAVEDYKHFSSECPFIPREAGEAYDIFPTSMDDPEQROFRALANOVVGMFVVDKLEN 120  
 |||||  
 OY 130 RIOELACSLIESLRPOGQCFTEDYAEPPIRIFMLAGLPREDIDHLKYLTDQMTRPDG 189  
 |||||  
 DB 121 RIOELACSLIESLRPOGQCFTEDYAEPPIRIFMLAGLPREDIDHLKYLTDQMTRPDG 180  
 |||||  
 OY 190 SMFAFAKEALDYLLPIITIOOROKPGT-----DAISIYANGOVNRP--ITSDEAKRM 241  
 |||||  
 DB 181 SMFAFAKEALDYLLPIITIOOROKPGNPNQDITDCLAMEKEKHNOSEPTTESLENT 240  
 |||||  
 OY 242 CGLLVGLDVTYVNFSLFSMEFLAKSPENROELIERPERI-----PA 283  
 |||||  
 DB 241 AVDLFGAGTETSTTLRYVALLLKLRPEVAKVQEEIERVIGNRSPQMDRSHMPTTDA 300  
 |||||  
 OY 284 ACEELLRRSLVADG--RLITSDEYFHGVOLKKGDOILL-----PQMLSG- 326  
 |||||

DB 301 VHEVORYTDLPTSLPHAVTCDIKFRNRYLIPKFTTILISTSVLHDNKEFPNEMEDPH 360  
 |||||  
 OY 327 --LDERENACPMHVDPSROKVSHTTFGSHLCLGQHLARRRITVTLKEMLTRIPDPSIA 384  
 |||||  
 DB 361 HFLDEGN-----FKKSKY-FMPFSACKRICVGEALAMELFLTLTILONFNKLSTLV 412  
 |||||  
 OY 385 PGAQIOHKSGIVSGVQALP 403  
 |||||  
 DB 413 DPRKND-TYPEVNGFASVP 430  
 |||||

RESULT 3  
 R60777  
 ID R60777 standard; Protein: 396 AA.  
 AC R60777:  
 DT 21-JUN-1995 (first entry)  
 DE Mycinamicin IV hydroxylating protein.  
 KW Mycinamicin; hydroxylase; macrolide; antibiotic; Micromonospora;  
 KM Micromonospora griseorubida.  
 OS Micromonospora griseorubida AT11725CN3.  
 PN J06253853-A.  
 PD 13-SEP-1994.  
 PF 09-MAR-1993; 047638.  
 PR 09-MAR-1993; JP-047638.  
 PA (ASAH ) ASAH KASEI KOGYO KK.  
 DR WPI: 94-328997/41.  
 DR N-PSDB: Q73674.  
 PT DNA encoding a protein having mycinamicin IV hydroxylating  
 activity - for prodn. of mycinamicin, a macrolide antibiotic  
 PS Claim 1: Page 12-14; 23pp; Japanese.  
 CC The amino acid sequence of a protein having mycinamicin IV hydroxylase  
 CC (MH) activity. The gene encodes a protein of 396 a.a. The DNA was  
 CC obtained from the macrolite-producing bacterium Micromonospora  
 CC griseorubida AT11725CN3/PT5507. The gene was isolated from the plasmid  
 CC PT5507. The protein encoded by this plasmid can be used to produce  
 CC mycinamicin IV in PT5507-deficient Micromonospora strains.  
 SQ Sequence 396 AA;

Query Match 16.7%; Score 365; DB 1; Length 396;  
 Best Local Similarity 28.5%; Pred. No. 1, 8e-28;  
 Matches 105; Conservative 57; Mismatches 168; Indels 38; Gaps 8;

OY 63 WINTRGQILREAYEDYRHFSSECPF-----PREAGEAYDIFPTSMDDPEQROFRA 113  
 |||||  
 DB 43 WLVTYR-----YEDYRAVLGQGRFVRCPSMTRDERTREPMKGGLLSMDDPEHRLRR 95  
 |||||  
 OY 114 LANOVVGMFVVDKLENRIOELACSLIESLRPOGQ-CNFTEDYAEPPIRIFMLAGLPREE 172  
 |||||  
 DB 96 LVVKAFTARAEELRRAREIAHELVDOMATGQPADLVAMFARQLPVRYICELGVP 155  
 |||||  
 OY 173 DIPHKLKYLTDQMTRPDGS-----TFAEAKELDYLLPIITIOOROKPRTDAISIV 223  
 |||||  
 DB 156 D-----HDFTRMSGALSTAEVTAEMEORAEQAVYMGDLIDRRKEPTDOLVSL 208  
 |||||  
 OY 224 ANGOVNGRPIITSDAKRMGGLLVGLDVTYVNFSLFSMEFLAKSPENROELIERPERIPA 283  
 |||||  
 DB 209 VQARQDDSLSEBELLDIAIGLVAGYESTTTQIAOFVLLMTRPLRRQGLDRPELIS 268  
 |||||  
 OY 284 ACEELLRRSL--VADGRILTSDEYFHGVOLKKGDOILLPQMLSGDLDERENACPMHVD 340  
 |||||  
 DB 269 AVEELTRWVPLGVTAFPRVAEDVTLRGVTTIACGPVLASTGANRDAQFPDADRIDV 328  
 |||||  
 OY 341 SRQKVSHTTFGSHLCLGQHLARRRITVTLKEMLTRIPDPSIA--PGAQIOHKSG-IVSG 398  
 |||||  
 DB 329 DRTPNQHLGFGHGVHLCAPLAREVLALEVLRLRPGIRLGPETQLRMSEGMLRG 388  
 |||||  
 OY 399 VQALPLVM 406  
 |||||  
 DB 389 PLELPLVM 396  
 |||||

RESULT 4

R77867  
ID R77867 standard; Protein: 408 AA.  
AC R77867;  
DT 13-NOV-1995 (first entry)  
DE S. clavuligerus ORE10 product.  
KW clavulanic acid; clavulinate; antibiotic; beta-lactamase-inhibitor.  
OS Streptomyces clavuligerus.  
PN CA2108113-A.  
PD 09-APR-1995.  
PF 08-OCT-1993; 108113.  
PR 08-OCT-1993; CA-108113.  
PA (UVAL-) UNIV ALBERTA.  
PI Alldoo KA, Jensen SE, Paradkar AS;  
DR WPI: 95-207301/28.  
N-PSDB: Q91580.  
PT Clavulanic acid biosynthesis enzymes and corresp. DNA - useful for  
PT biosynthesis of the antibiotic in Streptomyces hosts which do not  
PT naturally produce clavulanate  
PS Clam 32; Fig.19; 41pp; English.  
CC A 15 kb fragment S. clavuligerus NRRL 5741 genomic DNA (Q91580),  
CC extending downstream from pcbC, included 10 ORFs encoding the  
CC enzymes required for clavulinate biosynthesis. The ORE10  
CC product (R77867) showed high similarity to cytochrome P450-type  
CC enzymes from other Streptomyces spp.  
SQ Sequence 408 AA;

Query Match 15.2%; Score 332; DB 1; Length 408;  
Best Local Similarity 29.0%; Pred. No. 3.9e-25;  
Matches 122; Conservative 59; Mismatches 185; Indels 54; Gaps 18;

OY 16 PHEPHLVDFEDMYNSAGVQEAAMVLOESNVDLVWTRCNGH-WIATRGOLIREA 74  
DB 14 PAIRPMRRCVPD---PPRLQAGLRSGKASRYT-----IM---DGSQWLVYTHACARAV 62  
OY 75 YEDYRHFS-SECFIP-----BEAGEYDFIPTSMDDPEQROFRA-----LANQV 118  
DB 63 LGRFRPTAATVSAFGPMLRTSOLVIRANPEASFI--RMDDPHSRLSRLMTRDFLARRA 120  
OY 119 VGM-PPVVDKLENIQIOLACSLIESLRPOCQNTEDYAEFFPIRIMMLAGLPREDIPHL 177  
DB 121 EALRPVAVREL---LDLILGLVNGERR---VDLVAGLITTPVPSRYTLLFGADDREFFI 174  
OY 178 K---VLTDOMTPRPGSMTFAEAKELDYLIPIEORROKPGTDIAISYANGVNGRP 233  
DB 175 EDRSAVLIDRGTYPE---QVAKARDELIDGTLRLVEERIEENPCTDLISRLVITQVPRGHL 231  
OY 234 TSDAARMCGLLLVGSLDVVNFLSFMEFLAKSPENROELIERPERIPAAECCELLRFS 293  
DB 232 RVEEMVPMCRLLLVAGHGTTTSGASLSLSTLDRPELAGRTEDPALPRKAVELLRFHS 291  
OY 294 LVADG--RLITSYEFHGVQKKGDQILLPQMLSGIDERNACPMVDSRQVSHTTFG 351  
DB 292 IYONGIARRAVEDVDLDVLRAGEGVVLSLNGNDETVLPDRDVRDARRHLAFG 351  
OY 352 HGSHLGLGQHLAR---REIIVTKEWLTRIPDSIA--PGAQIOHKSGIYS--GVQALPLW 406  
DB 352 HGMHQLGOMLRAVELLEILAAVLRMM---PCARLAVPEELDFRREVSYSGLGALPVTW 408

RESULT 5  
ID W33274  
W33274 standard; Protein: 587 AA.  
AC W33274;  
DT 17-FEB-1998 (first entry)  
DE S. fradiae tylosin biosynthesis gene product tyIH.  
KW tylosin; biosynthetic gene product; production; antibiotic; tyIH.  
OS Streptomyces fradiae.  
PN US5672497-A.  
PD 30-SEP-1997.  
PF 21-DEC-1995; 575843.  
PR 12-MAY-1989; US-351350.  
PR 21-MAR-1986; US-842330.

PR 25-JUL-1986; US-890670.  
PR 24-FEB-1987; US-018237.  
PR 06-AUG-1991; US-742222.  
PR 28-JUL-1993; US-107232.  
PR 17-FEB-1994; US-198672.  
PR 21-DEC-1995; US-575843.  
PA (ELIL ) LILLY & CO ELI.  
PI Cox KL, Fishman SE, Hersherberger CL, Seno ET;  
DR WPI: 97-488860/45.  
N-PSDB: T58686.  
PT DNA encoding Streptomyces fradiae tylosin biosynthesis gene products  
PT - for increasing tylosin production in Streptomyces spp.  
PS Clam 20; Columns 21-38; 38pp; English.  
CC The present sequence is the Streptomyces fradiae tylosin  
CC biosynthetic gene product tyIH, useful to increase the production  
CC of the antibiotic tylosin in Streptomyces spp. .  
SQ Sequence 587 AA;

Query Match 15.2%; Score 330.5; DB 1; Length 587;  
Best Local Similarity 29.3%; Pred. No. 9.7e-25;  
Matches 105; Conservative 62; Mismatches 170; Indels 21; Gaps 11;

OY 63 WIATRGOLIREAYEDYR--HFSSECFIPREAGEA-YDFIPTSMDDPEQROFALANQV 118  
DB 143 WLISRODHVALLADPRVSIH-PAKLPRLSPSDGEAASRLTLDPDPRGALRGHPIPE 201  
OY 119 VGMPPVVDKLENIQIOLACSLIESLRPOG--QCNFTEDYAEFFPIRIMMLAGLPREDIPHL 177  
DB 202 FGIRRVRELSPSEQIVTGLDPLTGARGDEADLADFAFMATQVICRLDIPYERDYP 261  
OY 178 KYLTDOMTPRPGSMTFAEAKELDYLIPIEORROKPGTDI--SVANGVNGRPITS 235  
DB 262 QERTEDQTRPAAGEEALFELLDVLDRLISKTGCHESGDMLGMYA--QARGGGLSH 319  
OY 236 DEAKRMCGLLLVGSLDVVNFLSFMEFLAKSPENROELIERPERIPAAECCELLRFS 295  
DB 320 ADVLDNAVLLLAAGHETTASMTMSVULVLIQHPTAMRELIVNGLPGAVDELLRYLS-I 378  
OY 296 ADG--RLITSYEFHGVQKKGDQILLPQMLSGIDERNACPMVDSRQVSHTTFGHG 353  
DB 379 ADGIRSATVADIDIGITTTAGDGLVFLAANRDEAVSEPEAFDIHRSARHNVAFGYG 438  
OY 354 SHCLIGQHLARREIIVTKEWLTRIPDSIAP-----GAQIOHKSGIYSGVQALPLW 406  
DB 439 PHOCLGONLARMELEVALCAVLERLP--ALRPYTVAGLTKASDSA-VFQVYELPVAW 493

RESULT 6  
ID R11349  
R11349 standard; Protein: 406 AA.  
AC R11349;  
DT 05-JUN-1991 (first entry)  
DE Cytochrome enzyme P450SU1.  
KW Cytochrome P450; P450SU1; P450SU2; herbicide resistance.  
OS Streptomyces griseolus.  
PN M09103561-A.  
PD 21-MAR-1991.  
PF 27-AUG-1990; U04785.  
PR 11-SEP-1989; US-405605.  
PR 12-JAN-1990; US-464499.  
PR 23-AUG-1990; US-569781.  
PA (DUPQ ) DU PONT DE NEMOURS CO.  
PI Dean C, Harder PA, Leto KJ, Lichtner FT, Odell JT;  
PI O'Keefe DP, Omer CA, Komesser JA;  
DR WPI: 91-102077/14.  
N-PSDB: Q11126.  
PT DNA encoding cytochrome P450 enzymes - and electron donating  
PT iron-sulphur proteins, used to confer herbicide resistance to  
PT plants and microorganisms  
PS Clam 13; page 151; 224pp; English.  
CC This cytochrome P450 enzyme, P450SU1 is expressed alongside the iron  
CC sulphur protein P45-B, by a DNA sequence contained in a recombinant

CC plasmid. Host Streptomyces species are transformed with the plasmid  
CC and are used to coat a plant seed to transform the plant. The res-  
CC ultant transformants are resistant to herbicides.  
CC See also 011127.  
SQ Sequence 406 AA:

Query Match 15.1%; Score 330; DB 1; Length 406;  
Best Local Similarity 26.3%; Pred. No. 6,2e-25;  
Matches 104; Conservative 74; Mismatches 193; Indels 24; Gaps 11;

QY 31 PSLMSAGVQ--EAMAVLOESNVDLWTRKNGGH-WIATRGQLIREAVEDYR-----79  
D 17 PSMRSCPYDLDPDGYAQLRTPGLRHVTLVDGQAMVVTKEHAKRLIGDPRLSNRTDD 76  
QY 80 HFSSECFPIR--REAGEAYDFIPTSMPPRQORFALANOVMPVVDKLENTIOELACS 137  
D 77 NFPAVTPREAVAVESFOAF-----IGDPRPHGTRRKMTISEFTVKRIKGRPEVEEYVHG 132  
QY 138 LIESLRPGQ-QCNFTEDYAEPPPIRIFMLAGLPEEDIPHLKYLTDQMRPPDSMTFAEA 196  
D 133 FLEMLAAGPTADLVGSOFALPVPSSVTCRLGLGVADHEFFQDASKRLVSTDAQSLTA 192  
QY 197 KEALYLLPIIQOROKRGTDAI-SIVANGVNGRPITSDEAKRMCGLLVGGDLTVVN 255  
D 193 RNDLAGYLDGLITQFTEBPAGLVGALVADQLANGE-IDREELISTAMLLIGHETTAS 251  
QY 256 FLFSMEFLAKSPENHOELIERPERIPACEELLRRSL--VADGRILTSDFEFGVOLK 313  
D 252 MSLSYITLLDHEQYALRADRSILVPGAVELLRYLALADIAAGVATRADIVEGHILIR 311  
QY 314 KGDQILLPQMLSGLDERENACPMHVDPSROKVSHTTEGHSHLCLGOHLARREIIVLKE 373  
D 312 AGEGVIVVANSIARNDGTVEYEDPDALDIHRSARHHLAFEGVHQCLOGNLARLELEYLMA 371  
QY 374 WLTRIPDFSLA-PGAIOIHKSG-IVSGVALPLVW 406  
D 372 LMDRVPTLRILAVPEQDLVLRPCTTIGGVNELPVTW 406

RESULT 7  
W54389  
ID W54389 standard; Protein: 411 AA.  
AC W54389;  
DE 18-AUG-1998 (first entry)  
KW Actinomadura hibisca polyketide synthase protein 10.  
KW Multienzyme; infection; fungi; yeast; gram-positive bacteria; virus;  
OS dihydrobenzo(a)naphthacenequinone aglycon; antibiotic; pradimicin.  
OS Actinomadura hibisca.  
PN WO9811230-A1.  
PD 19-MAR-1998.  
PF 13-SEP-1996; U14791.  
PI 13-SEP-1996; WO-U14791.  
PA (BRIM) BRISTOL-MYERS SQUIBB CO.  
PI Dairi T, Oki T;  
DR WPI: 98-207391/18.  
DR N-PSDB: V26609.  
DR P-PSDB: W54389, W54389, W54390.  
PT Actinomadura polyketide synthase genes - useful for preparation of  
PT pradimicin  
PS Disclosure: Page 55-56; 71pp; English.  
CC The Actinomadura hibisca polyketide synthase proteins W54380-W54390 form  
CC a multienzyme complex. The enzyme is used for the biosynthesis of a  
CC dihydrobenzo(a)naphthacenequinone aglycon preferably a pradimicin which  
CC is an antibiotic useful against systemic fungal infections caused by  
CC Candida albicans, Aspergillus fumigatus and Cryptococcus neoformans. It  
CC is also active against a wide variety of fungi and yeasts, some  
CC Gram-positive bacteria and viruses.  
SQ Sequence 411 AA:

Query Match 15.1%; Score 329; DB 1; Length 411;  
Best Local Similarity 26.4%; Pred. No. 7.9e-25;

Matches 114; Conservative 61; Mismatches 199; Indels 58; Gaps 16;

QY 4 TIOSNMNLAPLRPHVEPHLVDFDWMNP-----SNLSAGVQEAAMAVLOESNVDLWTRCN 59  
D 9 TVDPREDVTPAPFFRPD-----DPFQPCENARLAKSDPAKAVLP-----F 50  
QY 60 GGH-WIATRGQLIREAVEDYRHFSSSEC---PFIPREAGEAYDFIPTSMPPRQORFALANOV 107  
D 51 GDHAAVVTTRADVRAEFTVSD-RRFSKAVTRPGAPR-----LIPQORSKSLVMDPPE 102  
QY 108 QROFRALANOVMPVVDKLENTIOELACSLESIRPGQ-QCNFTEDYAEPPPIRIFMLL 166  
D 103 HTRMRKIVSAFTARRVEGKRAHVRLDTSQFVDEWHEGPPADLIHALLPLPVTVICEM 162  
QY 167 AGLPEEDIPHLKYLTDQMRPPDSMTFAEAK-----ALYDLIPIIQOROKRGTDAI 220  
D 163 LGVPEDRRFRFQWTTBRL-TIGAPALAOADELKAANGRLGLAELIDKTRAPADLL 221  
QY 221 SIVANGVNGRPITSDEAKRMCGLLVGGDLTVVNFLFSMEFLAKSPENHOELIERPER 280  
D 222 SLISRAHADQ-GLSEBELLTFCGWTLLAGYHTTTAITSVYHLREPSRYARLRDPDG 280  
QY 281 IPACEELLRRSLVADG-----RILTSDFEFGVOLKKGQDILLPQMLSGLDERENACPM 336  
D 281 IPAAVEELL-RGQIIGGAGAIRIAVEDVEVGGLTVRAGEAVIPLPNAARPEVADPE 339  
QY 337 HVDPSROKVSHTTEGHSHLCLGOHLARREIIVLKEWILTRIPDFSLA-PGAIOIHKSGI 395  
D 340 ELDLGRITDNRHIALGCHITICLGAFLARLELOVYLTVERPALRIALIDDADITWRPGL 399  
QY 396 V-SGVQALPLVW 406  
D 400 AFAPRDALPIAW 411

RESULT 8  
W11585  
ID W11585 standard; Protein: 398 AA.  
AC W11585;  
DE 02-APR-1997 (first entry)  
KW Streptomyces pristinaespiralis SnpF gene product.  
KW Streptogramin B; antibiotic; biosynthesis; pristinamycin;  
KW virginiamycin; pipicolinic acid; cyclodeamination; papa; snbA; snbF;  
KW pipA; 3-hydroxypicolinic acid; hydroxylation.  
OS Streptomyces pristinaespiralis.  
PN WO9601901-A1.  
PD 25-JAN-1996.  
PF 04-JUL-1995; F00889.  
PI 08-JUL-1994; FR-008478.  
PA (RHON) RHONE-POULENC RORER SA.  
PI Barriere JC, Blanc V, Blanche F, Crouzet J;  
PI Debussche L, Paris JM, Thibaut D, Bamas-Jacques N;  
PI Dutruc-Rosset G, Famechon A;  
DR WPI: 96-097631/10.  
DR N-PSDB: T58535.  
PT New streptogramin B derivs. useful as antibiotics - produced by new  
PT mutants of Streptomyces having altered genes for streptogramin B  
PT biosynthesis  
PS Example 1; Page 113-114; 146pp; French.  
CC The papa gene of S.pristinaespiralis is involved in the biosynthesis  
CC of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for  
CC pristinamycin IA. Upstream of the papa gene, on the complementary  
CC strand, is the snbA gene coding for 3-hydroxypicolinic acid-AMP ligase.  
CC The region between these two genes was sequenced and two open reading  
CC frames were identified. The first (pipA) decodes to an amino acid  
CC sequence with homology to ornithine cyclodeaminase from Agrobacterium  
CC tumefaciens. The pipA gene product is likely to catalyse the cyclo-  
CC deamination of lysine, leading to production of pipicolinic acid.  
CC Mutations in the pipA gene were shown to affect pipicolinic acid  
CC synthesis but not the synthesis of 3-hydroxypicolinic acid. The second  
CC open reading frame (snbF) could be decoded to give a product with  
CC homology to hydroxylases of the cytochrome P450 type.  
CC Disruption of the pipA and snbF genes can be used to produce



CC strains of *S.pristinaespiralis* which are unable to produce the  
CC antibiotic pristinamycin I but which may be able to produce new,  
CC modified forms of it.  
SO Sequence 398 AA;

Query Match 14.9%; Score 324.5; DB 1; Length 398;  
Best Local Similarity 29.0%; Pred. No. 2.1e-24;  
Matches 106; Conservative 59; Mismatches 159; Indels 41; Gaps 12;

QY 74 AYEDYRA-----FSSECPFIPREAGYDFIPISMPDEOROFALANQVGM 121  
DB 36 AFHFHRAVDLVTAADPGVYSSQSLRSPGSQALSEQILSYIDPPMHTLRRLVSOAFTP 95  
QY 122 PVYDKLENIROELACSLIESLRPGQC-NFTEDYAEPPPIRIFMLAGLREEDIPHLKYL 180  
DB 96 RYVADLEPRVTELAGQLLDVA--DGTFFDLVADRPYPLVIVIAELLGVPRADTLFRSW 153  
QY 181 TDQWTR-----PDGSMTEFAKEALYDYLPIIEORROKPGTDAISIVA 224  
DB 154 SDRMLQVQVADPADMQGDDADEDYQRLVKEPMAMHAYLHDHTDRARRANDLISALV 213  
QY 225 NGOVNRPITSDAKKRMCGLLVGLDTVVNFLSFSMEFLAKSPENROELTERPER--IP 282  
DB 214 AARVEGERLTDEQIVFERGALLMAGHVSTSMLLGNFVLCIKDHP--RAEAAARADRSLLP 271  
QY 283 AACCELLR-RFSLVADGRITLSDYEFHGVLKKGDQILLPQMLS-GLDERENACPMHVD 340  
DB 272 ALIEEYLRPRPIVMAVTTKTQVLAGTTIPAG-RMNVPSLSANHDEQVFTDPDLHDL 330  
QY 341 SROKVSHTTGHGSHLCLGQHLARREIYTLKEWLTIRIPDSIAPGAQIQ-HKSGIVSGV 399  
DB 331 ARBG-RQIARGHGHYGLGAPLANLEGRALDELFDRFPPSPIDGAKLRHRCGLF-GV 388  
QY 400 QALPL 404  
DB 389 KNEPL 393

RESULT 9  
ID R11350 standard; Protein; 403 AA.  
AC R11350;  
DT 05-JUN-1991 (first entry)  
DE Cytochrome enzyme P450SU2.  
KW Cytochrome P450; P450SU1; P450SU2; herbicide resistance.  
OS Streptomyces griseolus.  
PN W09103561-A.  
PD 21-MAR-1991.  
PF 27-AUG-1990; U04785.  
PR 11-SEP-1989; US-405605.  
PR 12-JAN-1990; US-464499.  
PR 23-AUG-1990; US-569781.  
PA (DUPQ ) DU PONT DE NEMOURS CO.  
PI Dean C, Harder PA, Leto KJ, Lichtner FT, Odell JT;  
PI O'Keefe DP, Omer CA, Romesser JA;  
DR WPI: 91-102077/14.  
DR N-PSDB: Q11127.  
PT DNA encoding cytochrome P450 enzymes - and electron donating  
PT iron-sulphur proteins, used to confer herbicide resistance to  
PT plants and microorganisms  
PS claim 15; page 158; 234pp; English.  
CC This cytochrome P450 enzyme, P450SU2 is expressed alongside the iron  
CC sulphur protein FeS-A, by a DNA sequence contained in a recombinant  
CC plasmid. Host Streptomyces species are transformed with the plasmid  
CC and are used to coat a plant seed to transform the plant. The res-  
CC ultant transformants are resistant to herbicides.  
CC See also Q11126.  
SO Sequence 403 AA;

Query Match 14.8%; Score 322; DB 1; Length 403;  
Best Local Similarity 28.4%; Pred. No. 3.9e-24;

Matches 103; Conservative 64; Mismatches 166; Indels 30; Gaps 13;

QY 63 WIAFRGQIREAVDYVHFSESC-----PIPRRAGAYFIPIR--SMDPEOROFALA 115  
DB 52 WLTVRHODVRAVLGDDPR-FSADAHRTGPPPLTAGRGRICTNPFLRMDDENARLRML 110  
QY 116 NOVGMFVVDKLENIROELACSLIESLRP-QGCFNFTEDYAEPPPIRIFMLAGLPEEDI 174  
DB 111 TADRIYAKVEAMREYORLADDDVDRMTTGTSADYTFEALPLPSLVICLLGVPREDH 170  
QY 175 PHL---KYLTDQWTRPDGSMTEFAKEALYDYLPIIEORROKPGTDAIS-IVANQVN 229  
DB 171 AFEQERSRVLTLTRSTPE---EVRAADELLELYLARLARTRKRRPDDAISRLVARGEL 227  
QY 230 GRPITSDEAKRMCGLLVGLDTVVNFLSFSMEFLASPEHROELTERPERIPACCELL 289  
DB 228 DTQIAT-----MGRLLVAVGGEGLDDPQIATVMGILLRNPDOLARAEPAVLCVABEL 282  
QY 290 RRFSLVADG--RILTSDEYEFHGVLKKGDQILLPQMLSGDERENACP--MHVDSROKV 345  
DB 283 RYLITVINGVPRIATDEVLLIGRTIAGEGVLC--MISSANROAEVPPGGDDLDVADAR 340  
QY 346 SHTTFGHSHLCLGQHLARREIYTLKEWLTIRIPDSIA-PGAQIQHKSGL-VSGVALP 403  
DB 341 RHVAFGSGVHOCLOGLPRLAVELQIAETLRLRLPDLRLAVPHEIIPRGDAIYGVSLP 400  
QY 404 LVW 406  
DB 401 IAW 403

RESULT 10  
ID R38309 standard; Protein; 412 AA.  
AC R38309;  
DT 04-DEC-1993 (first entry)  
DE Sequence of the P450-soy protein.  
KW P450soy; soyC gene; oxidation; haem protein.  
OS Streptomyces griseus ATCC 13273.  
PN W09312236-A.  
PD 24-JUN-1993.  
PF 16-DEC-1992; U10885.  
PR 16-DEC-1991; US-807001.  
PA (DUPQ ) DU PONT DE NEMOURS & CO E I.  
PI Omer CA, Sariaslani FS, Trower MK;  
DR WPI: 93-214178/26.  
DR N-PSDB: Q45569.  
PT Constitutive expression of P450 SOY (SOYC) and ferredoxin soy  
PT (soyD) in Streptomyces - used for oxid. of organic chemicals  
PS Example; Figure 2; 45pp; English.  
CC Cytochrome P450soy was purified from *S. griseus* ATCC 13273. Two  
CC similar forms of P450soy were isolated. P450soy-delta, is derived  
CC from P450soy by in vitro proteolysis during isolation. One of the  
CC tryptic peptide fragments of cytochrome P450soy and of of the  
CC P450soy-delta protein were subjected to automated degradation. The  
CC NH2 terminal sequences are given in R38306 and R38307. A mixture of  
CC oligos that consist of possible DNA sequences that could encode the  
CC Aas FeVHQCIL of the tryptic peptide was made. It consists of the  
CC sequences in Q43290-23. The oligo mixture was end-labeled and used  
CC to probe the EMBL4 library of *S. griseus* DNA. Hybridising plaques  
CC were isolated and a 4.8kb SacI DNA fragment was isolated from one  
CC clone that hybridised to the oligo probe mixture. As segment of the  
CC 4.8kb fragment was sequenced and found to contain an ORF. Within  
CC this ORF was a section that corresp. exactly to the AA sequence  
CC determined from the cytochrome P450soy tryptic peptide (see Q45569,  
CC R38309). The gene encoding the P450soy protein was called soyC. Five  
CC nucleotides downstream from the stop codon for soyC another ORF was  
CC identified. This ORF encodes an apparent ferredoxin-like protein.  
CC The gene was designated soyB and the protein ferredoxin-soy.  
SO Sequence 412 AA;

Query Match 14.5%; Score 316; DB 1; Length 412;





KW Daunomycin C-14 hydroxylase; doxa gene; doxorubicin; daunomycin;  
KW 13-dihydrocarminomycin; carminomycin; anthracycline;  
KW anticancer; cytostatic; cancer; therapy; plasmid pANT199.  
OS Chimeric - Streptomyces sp. strain C5.  
FH Chimeric - synthetic.  
FH Key location/Qualifiers  
FT Cleavage-site 31  
FT Protein /note="enterokinase cleavage site"  
FT 53..474  
PN MO974439-A2. /note="native daunomycin C-14 hydroxylase"  
PD 27-NOV-1997.  
PF 22-MAY-1997; 008690.  
PR 24-MAY-1996; US-653650.  
PA (OHIS ) UNIV OHIO STATE RES FOUND.  
PI Desanti CL, Dickens ML, Strohl WA;  
DR WPI: 98-018495/02.  
DR N-PSDB: V01452.  
PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase  
PT - also hydroxylation and oxidation of other anthracycline(s) with  
PT the same enzyme  
PS Disclosure: Fig 9: 59pp: English.  
CC This protein comprises an N-terminal modified enzyme derived from  
CC the daunomycin C-14 hydroxylase (see W36128) of Streptomyces sp.  
CC strain C5. It is encoded by a gene construct (see V01452) in  
CC plasmid pANT199. In this plasmid, the doxa gene (see V01447) is  
CC translationally fused with a leader sequence encoding 6  
CC histidine residues so that the fusion protein can be affinity  
CC purified on a nickel-agarose gel. Daunomycin C-14 hydroxylase  
CC is a P450-like enzyme capable of converting daunomycin to the  
CC anticancer agent doxorubicin. Host cells, especially Streptomyces  
CC host cells, transformed with pANT199 can be used in methods for  
CC the production of doxorubicin from daunomycin or for the  
CC hydroxylation and oxidation of other anthracyclines.  
SQ Sequence 474 AA;

Query Match 7.4%; Score 161; DB 1; Length 474;  
Best Local Similarity 23.6%; Pred. No. 7.9e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 QEAMNVLOESNVLDVWTRNGGHWITRGOLIREATEDYRHSSF--CPFIREAGEAY 96  
DB 83 REAGPVV-EVNAP-----AGGPAWYTTDDALREVLADRFVKGPDLPATAMRGVDDGL 135  
QY 97 D-----PIPTMDPEEOROFALANQVGMVVDKLENRIOELACSLIFSL-----R 143  
DB 136 DIVPELRPTLLAVDGEDHRLRLRIHAPAFNFRRLAERTDRIATAIDRLLELADSSDR 195  
QY 144 PQGQCNFTEDYAEFPFIRIFMLAGLP-----EEDIPHLKYL-----TDQMTRPD 188  
DB 196 SGEPAEILIGFAVHFPFLVLCGLGVPTDPAMAREAVGVKALGLCGPQSGAGDGTDP 255  
QY 189 GSKTFPAEAKALDYLPITTEOROKGTDAISIVANGOVNGRITSDEAKRMCGLLVIG 248  
DB 256 GDVPDTSALSLSL--LLEAVHAARRKDTRTMTRVLYERPAQFEFGVSDDQLVYMITGLIFA 313  
QY 249 GLDPTVNVFLSFMSMEFLAKSPENHROELIERPERIPRACEELLR-----FSLVADGRILTS 303  
DB 314 GHDTTGSFLGF--LAEVLAGRLADADGDAISRVEEALRHNPVRYSL--WRPAAT 367  
QY 304 DYEFHGVQLKKGDQILLPOMLSGLDERENACPMHVDFSRQKVSHTTFFGHGSHLCLGQHILA 363  
DB 368 EVVIRGVRLPRGARVLYDIGETWDCRHHNDAPAHFHPDRSRRLRTFGDGRHVCIGBQLA 427  
QY 364 RRELIYTLKEMLTRIPDESTA-FGAQIQ--HKSGIVSGVQALPLVW 406  
DB 428 QLESRTMIGVLRSPQARLAVPEELRMCRCRKGATARTLDP-VW 472

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 06:41:03 ; Search time 66.92 Seconds

(without alignments)  
103.693 Million cell updates/sec

Title: US-09-246-451-2

Perfect score: 2160  
Sequence: 1 TTTETQSNAIPLPPHVE.....IVSGVQALPLVWDPAITTKAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 16761186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued\_Patents\_AA:\*  
2: /cgnl\_7/ptodata/1/iaa/5A\_COMB.pep:\*  
3: /cgnl\_7/ptodata/1/iaa/5B\_COMB.pep:\*  
4: /cgnl\_7/ptodata/1/iaa/6\_COMB.pep:\*  
5: /cgnl\_7/ptodata/1/iaa/PCITUS\_COMB.pep:\*  
6: /cgnl\_7/ptodata/1/iaa/Backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	363.5	16.7	416	3	US-09-320-878-18
2	340	15.6	403	5	5212296-9
3	330	15.1	406	5	5212296-6
4	316	14.5	412	1	US-08-102-863-11
5	316	14.5	412	4	PCR-US92-10885-11
6	271.5	12.5	419	3	US-09-335-409-8
7	161	7.4	422	2	US-09-096-982-5
8	161	7.4	422	2	US-08-653-650A-5
9	161	7.4	474	2	US-09-096-982-8
10	161	7.4	474	2	US-08-653-650A-8
11	158	7.2	443	2	US-09-096-982-9
12	158	7.2	443	2	US-08-653-650A-9
13	156	7.2	422	1	US-08-396-218-2
14	156	7.2	422	1	US-08-760-116-2
15	129	5.9	512	2	US-08-194-981E-5
16	123.5	5.7	512	3	US-09-320-878-7
17	118	5.4	513	3	US-08-948-564-6
18	115.5	5.3	490	1	US-08-201-118-7
19	115.5	5.3	490	2	US-08-238-821B-7
20	115.5	5.3	490	4	PCR-US95-05744-7
21	113.5	5.2	490	1	US-08-201-118-3
22	113.5	5.2	490	1	US-08-201-118-9
23	113.5	5.2	490	2	US-08-238-821B-3
24	113.5	5.2	490	2	US-08-238-821B-9
25	113.5	5.2	490	4	PCR-US95-05744-3
26	113.5	5.2	490	4	PCR-US95-05744-9
27	113	5.2	501	3	US-08-906-791-2
28	111.5	5.1	504	1	US-08-457-274A-25

29	111.5	5.1	504	4	PCR-US95-05758-25	Sequence 25, App1
30	110	5.0	492	3	US-08-724-466B-2	Sequence 2, App1
31	106.5	4.9	466	1	US-08-313-075A-50	Sequence 50, App1
32	102.5	4.7	472	2	US-08-622-166A-2	Sequence 2, App1
33	102.5	4.7	472	2	US-08-622-166A-4	Sequence 4, App1
34	102	4.7	497	3	US-08-724-466B-4	Sequence 4, App1
35	99.5	4.6	490	1	US-08-201-118-1	Sequence 1, App1
36	99.5	4.6	490	2	US-08-238-821B-1	Sequence 1, App1
37	99.5	4.6	490	1	PCR-US95-05744-1	Sequence 1, App1
38	99.5	4.6	498	1	US-08-457-274A-24	Sequence 24, App1
39	99.5	4.6	498	4	PCR-US95-05758-24	Sequence 24, App1
40	97.5	4.5	500	2	US-08-314-601-2	Sequence 2, App1
41	97.5	4.5	500	4	PCR-US95-13051-2	Sequence 2, App1
42	94.5	4.3	523	3	US-08-606-505B-67	Sequence 67, App1
43	94.5	4.3	1026	1	US-07-998-003A-95	Sequence 95, App1
44	94.5	4.3	1026	1	US-08-453-274B-95	Sequence 95, App1
45	94.5	4.3	1026	1	US-08-453-695A-95	Sequence 95, App1

#### ALIGNMENTS

```
RESULT 1
US-09-320-878-18
; Sequence 18, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320, 878A
; EARLIER FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141, 908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073, 538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846, 247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119, 139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100, 880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087, 080
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-18
```

Query Match 16.7% Score 363.5; DB 3; Length 416;  
Best Local Similarity 28.3%; Pred. No. 1.3e-29;  
Matches 97; Conservative 61; Mismatches 164; Indels 21; Gaps 7;

```
QY 76 EDVRRHSSCPPIIPREGEAYDFIPSSMDPPEQROPALANOVVGMVYVLEKRIQELA 135
DB 72 KDMR--NSTPTPLEAEALAHNMHIES--DPRHTRLKLVAREFTMRVLELPPROVEIV 127
QY 136 CSLIEEL--RPQOCNTEDEYAEFPPIRIFMLAGLEEDIPUKXYLTDOMTQPPDSMTF 193
DB 128 DGLVDAMLAAPDGRADLMESLAMPPLPTIVISELGVPEPDRAPAFVWTDQAFVPPDDPAQ 187
QY 194 AEKKEALVDLPIITIEQRQKPGTDAIS--IVANGVNGRITISDEAKRKMGGLLVGLDT 252
DB 188 QTMAEMSGYLSRLIDSKRGQDEDDLSALVTRSDSGSRUTSEELGMHILLVAGHET 247
```

```

QY 253 VVNFSLFSMEFLAKSPENHROELIERPERIPACCELLRRFSLVADGRILTSYEF----- 307
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 248 IVNLIANGVYALLSHPDJALALRADMTLLDGAVEEMLR-----YEGPVSATYRFVEYV 302
QY 308 --HGVOUKGDOILLPOMLSGLDERENACPMHVDPSRQVSHTTGSHSLCIGHLARR 365
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 303 DLDGTVIPAGDVTLYVLADAHRTPERFDPHREDIRDRDTAGHIAFGHIFCIGABLARL 362
QY 366 EIVVTLKEMLTRIPDS--IAPGAOIHKSGIYSGVOALPLW 406
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 363 EARIAVALLERCPDLALDVSPEELVWYPMTRIGLALPIRW 405

RESULT 2
5212296-9
; Patent No. 5212296
; APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
; J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.
; TEPPERMAN, JAMES M.
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
; CYTOCHROMES
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/569,781
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 464,499
; FILING DATE: 12-JAN-1990
; APPLICATION NUMBER: 405,605
; FILING DATE: 11-SEP-1989
; SEQ ID NO:9:
; LENGTH: 403
5212296-9

```

```

Query Match 15.6%; Score 340; DB 5; Length 403;
Best Local Similarity 28.9%; Pred. No. 3.5e-27;
Matches 105; Conservative 63; Mismatches 165; Indels 30; Gaps 13;

```

```

QY 63 WIAIRGQLIREAVEDYRHRSSSEC-----PIPREAGEAYDFT--SMDPPEQROFRALA 115
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 52 WATRHQDVRAYLVGDR-PSADAHRTGFPPLTAGREIIGTTLRMDDPEHARLRRL 110
QY 116 NOVVGHPVVDKLENRIQELACSLIESLRP-QGQCNFTEDYAEPPRIIFMLLAGLPEEDI 174
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 111 TADFIYKKVYEAHPREYQALADLVDMTGRISADLVTERALPLPSLVICHLIGVYEDH 170
QY 175 PHL---KYLTDQMPRPGDSMTFAEAKALYDLPRIEQRQKPGTDAIS-IVANGOVN 229
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 171 AFGQERSRVLLTFRSTPE---EYVRAQDELLLEYLARLARTKRRERPDATISRLVARGELD 227
QY 230 GRITSDDEAKRMGCLLVGGLDVVNFSLFSMEFLAKSPENHROELIERPERIPACCELL 289
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 228 DTQIAT----MGRLLLVGHETTAMTALSTLVLLRNPDOLARLAEPALVGAVEELL 282
QY 290 RRESLVADG--RLTSDYEFHGVLKKGDOILLPOMLSGLDERENACP--MHYDPSRQV 345
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 283 RYLTITHNKVPRIATEDVILGRTIAGEGVLC--MISSANRAEYVPGDDLDVARDAR 340
QY 346 SHTTFGHSHLIGOHILARREIIVTLKEMLTRIPDSIA-PGAOIHKSGI-VSGVOALP 403
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 341 RHVAFGEFGVHOGCGLPRLAREIQIALETLLRLRPDLRLAVPHREIFPRGMATYGVHSLP 400
QY 404 LVM 406
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 401 IAM 403

```

```

RESULT 3
5212296-6
; Patent No. 5212296
; APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
; J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.

```

```

; TEPPERMAN, JAMES M.
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
; CYTOCHROMES
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/569,781
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 464,499
; FILING DATE: 12-JAN-1990
; APPLICATION NUMBER: 405,605
; FILING DATE: 11-SEP-1989
; SEQ ID NO:6:
; LENGTH: 406
5212296-6

```

```

Query Match 15.1%; Score 330; DB 5; Length 406;
Best Local Similarity 26.3%; Pred. No. 3.9e-26;
Matches 104; Conservative 74; Mismatches 193; Indels 24; Gaps 11;

```

```

QY 31 PSNLSAGVO--EAMAVLOSNNPDLWTRCNGH-WIATRGQLIREAVEDYR----- 79
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 17 PSNRSCPQYLPDQYQALRDTGPRLHRVTLYDGRQAVVTKHEAARLLGDPRLSNRDTD 76
QY 80 HFSSECPFIP--REAGEAYDFTSMDPPEQROFRALANQVGMVVDKLENRIQELACS 137
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 77 NFPAISPREAVRESQAF-----IGLDPEHGRTRRMTTSEFTVKRIKQMRVEVEYVNG 132
QY 138 LIESLRPG-QCNFTEDYAEPPRIIFMLLAGLPEEDIPHLKYLTQDMTRPGDSMTFAEA 196
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 133 FLEMLAAGPTADLVQSQALPVSWVICRLIGVYADHDFQDASKRLVQSTDAQALTA 192
QY 197 KEALYVLPRIEQRQKPGTDAI-SIVANGOVNRPRTSDEAKRMGCLLVGGLDTVYN 255
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 193 RNDLAGYLDGLTFQTEGCGAGLVGALVADQLANGE-IDRELISTAMLLLAGHETTAS 251
QY 256 FLFSMEFLAKSPENHROELIERPERIPACCELLRRFSL--VADGRILTSDFEFHGVOUK 313
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 252 MNSLSVITLLDHPDEQYALRADSLVPGAVEELRYLAIADIAGGVATADIEVECHLIR 311
QY 314 KGDQILLPOMLSGLDERENACPMHVDPSRQVSHTTGSHSLCIGOHILARREIIVTLKE 373
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 312 AGEVIVVNSIANRQDGTVEDPDALDIHRSARHHLAFGFGVHOGCIGOHILARELEVITLMA 371
QY 374 WILTRIPDSIA-PGAOIHKSGI-YSGVOALPLW 406
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 372 LMDRVPTRLAVPVEQLVLRPGTTIOGVNELPYTW 406

```

```

RESULT 4
US-08-102-863-11
; Sequence 11, Application US/08102863
; Patent No. 5466590
; GENERAL INFORMATION:
; APPLICANT: SARIASLANI, SIMA
; TITLE OF INVENTION: CONSTITUTIVE
; TITLE OF INVENTION: EXPRESSION OF P450SOY
; TITLE OF INVENTION: AND FERREDOXIN-SOY IN
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; ADDRESSER: AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```



; ZIP: 19898

```

; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patattin Ver 3.0

```

```

; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patattin Ver 3.0

```





```

; APPLICATION NUMBER: US/08/653.650A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22727/00131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 216-622-8458
; TELEFAX: 216-241-0816
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-653-650A-8

Query Match          7.4%; Score 161; DB 2; Length 474;
Best Local Similarity 23.6%; Pred. No. 2.2e-08;
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 OEAMAVLOESNPDLVWTRCNGHMIATRGQLIREAYEDYRHFSS--CPFIPTREAGEAY 96
DB 83 REAGPVV-EVNAP-----AGSPAWITDDALAREVLADPRVKGPDLPATMRKGVDDGL 135
QY 97 D-----FIFTSMDPREQROFRALANOVGMVVDKLENIQELACSLIESL-----R 143
DB 136 DIPVELRPFTLIAVDGEHRLRLRIHAPFNPRLAERTDRIAAIDRLTELADSSDR 195
QY 144 POGQCNFTEDVAPERRIRIFMLLAGLP-----BEDIPHLKYL-----TDQMTRPD 188
DB 196 SEPAELLIGGFAYHNPFLVLCGLVPTDPMAREAVGVALKALIGGQSGAGGDTDA 255
QY 189 GGMTFAEAKAELADYLPIIEQRORPGTDALSIYANGVNGRPITSDAKRMCGILLVG 248
DB 256 GVPPTSALESU--LLEAVHARRKDTRTMTVLYERAQAEFGVSDDQLVYMITGLIFA 313
QY 249 GIDTVVNFISFMEFLANSPEHROELIERPERIPACCELLRR-----FSLVADGRILTS 303
DB 314 GHDITGSGIFG---LLAEVLAGRLAADADGDAISRFEVALRHNHPVPTSL--WRFAT 367
QY 304 DYEHGVOLKKGDQILLPQMSGLDERENACPMHYDFSRQKVSHTTFGHSILCIGOHILA 363
DB 368 EVYIGVRLPRGAPVLYDIEGTNTDGRHNDAPHAPRPSRRRLTFGDPHYCIGEQOLA 427
QY 364 RELIIVTLKEMLTIRIPDSIA-PGAQIO--HKSGIVSGVALPLVW 406
DB 428 QLESRTMIGVLSRFPQARLAVPYEELRWCKRGAQTARLTLDLP-VW 472

RESULT 11
US-09-096-982-9
; Sequence 9, Application US/09096982
; Patent No. 5962293
; GENERAL INFORMATION:
; APPLICANT: Strohl, William R.
; APPLICANT: Dickens, Michael L.
; APPLICANT: Desanti, Charles L.
; TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CALFEI, HALTER & GRISWOLD
; STREET: 800 Superior Avenue, Suite 1400
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: USA
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/096.982
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22727/00131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 216-622-8458
; TELEFAX: 216-241-0816
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-096-982-9

Query Match          7.2%; Score 158; DB 2; Length 443;
Best Local Similarity 23.6%; Pred. No. 4.1e-08;
Matches 95; Conservative 50; Mismatches 210; Indels 48; Gaps 13;

QY 39 OEAMAVLOESNPDLVWTRCNGHMIATRGQLIREAYEDYRHFSS--CPFIPTREAGEAY 96
DB 52 REAGPVV-EVNAP-----AGSPAWITDDALAREVLADPRVKGPDLPATMRKGVDDGL 104
QY 97 D-----FIFTSMDPREQROFRALANOVGMVVDKLENIQELACSLIESL-----R 143
DB 105 DIPVELRPFTLIAVDGEHRLRLRIHAPFNPRLAERTDRIAAIDRLTELADSSDR 164
QY 144 POGQCNFTEDVAPERRIRIFMLLAGLP-----BEDIPHLKYL-----TDQMTRPD 188
DB 165 SEPAELLIGGFAYHNPFLVLCGLVPTDPMAREAVGVALKALIGGQSGAGGDTDA 224
QY 189 GGMTFAEAKAELADYLPIIEQRORPGTDALSIYANGVNGRPITSDAKRMCGILLVG 248
DB 225 GVPPTSALESU--LLEAVHARRKDTRTMTVLYERAQAEFGVSDDQLVYMITGLIFA 282
QY 249 GIDTVVNFISFMEFLANSPEHROELIERPERIPACCELLRRFSLV--ADGRILTS DYE 306
DB 283 GHDITGSGIFG---LLAEVLAGRLAADADGDAISRFEVALRHNHPVPTSLWRFATEV 339
QY 307 FHGVOLKKGDQILLPQMSGLDERENACPMHYDFSRQKVSHTTFGHSILCIGOHILARE 366
DB 340 INGVRLPRGAPVLYDIEGTNTDGRHNDAPHAPRPSRRRLTFGDPHYCIGEQOLA 399
QY 367 ILVTLKEMLTIRIPDSIA-PGAQIO--HKSGIVSGVALPLVW 406
DB 400 SRTMIGVLSRFPQARLAVPYEELRWCKRGAQTARLTLDLP-VW 441

RESULT 12
US-08-653-650A-9
; Sequence 9, Application US/08653650A
; Patent No. 5976830
; GENERAL INFORMATION:
; APPLICANT: Strohl, William R.
; APPLICANT: Dickens, Michael L.
; APPLICANT: Desanti, Charles L.
; TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CALFEI, HALTER & GRISWOLD
; STREET: 800 Superior Avenue, Suite 1400
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: USA
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/653,650A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gollick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22727/00131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 216-622-8458
; TELEFAX: 216-241-0816
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-653-650A-9
```

```

Query Match          7.2%; Score 158; DB 2; Length 443;
Best Local Similarity 23.6%; Pred. No. 4.1e-08;
Matches 95; Conservative 50; Mismatches 210; Indels 48; Gaps 13;
```

```

QY 39 QEAMAVLQESNVDLVWTRCNGHMIATRGQLIREAVEDYRHSSE--CPFIPEGAEAY 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 52 REAGPVV-EVNAP-----AGPRAMVITDDALAREVLADRFVKDPLATFAMRGVDGL 104
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 97 D-----PIPTSMDEPREROFALANQVGVYDKLENRIQELACSLIESL-----R 143
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 105 DIVEPELRPTLLIADGEDRRRLRIHAPAFNPRLAERTDRIAAIDLRLTELAUSSDR 164
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 144 POGCNETEDYAEPPRIPIFMLAGLP-----EEDIPHLKYL-----TDQWTRPD 188
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 165 SGPRAELIGGFAYHPRFLVLCGLGVPTDPMAREAVGLKALGLGPGSAGDGDIPA 224
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 189 GSWTFAEAKALYDYLPIIEQRQKPGTDAISIVANGQVNGRPTSDEAKRMCGILLVG 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 225 GDVPDTSLESL--LLEAVHARRKDTRTMTVLYERAQAEFGVSDDQLVVMITGLIFA 282
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 249 GLDTVYNFLFSMEFLAKSEHROELIERPERIPACEELLRFSLV--ADGRLISDYE 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 283 GHDTTGSFLGF--LLEAVLAGRLAADADDAISRFYEELRHHRPVYTLWRFPAATEVY 339
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 307 FHGVOLKKGDQILLPQMLSGLDERENACPMHVDPSHOKVSHTTFGHSHLCLGHLARRE 366
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 340 IRKVRILPRKAPVLDIEGTDTGRIHNDAPHAHFPRDRSRRLTFGDDPHYCTIGEOULAQLE 399
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 367 IIVTLKEMLTRIPDFSIA--PGAQIQ--HKSGLVSGVALPLVW 406
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 400 SRTMIGVLSRFPQARLAVPYEELRMCRCGAQTARLTDLF-VW 441
```

```

RESULT 13
US-08-396-218-2
; Sequence 2, Application US/08396218
; Patent No. 5695966
; GENERAL INFORMATION:
; APPLICANT: INVENTI, Augusto
; APPLICANT: BREME, Umberto
; APPLICANT: COLOMBO, Anna L.
; APPLICANT: HUTCHINSON, Charles R
; APPLICANT: OTTEN, Sharee
; APPLICANT: SCOTTI, Claudio
; TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAIKO, MARCELSTEIN, MURRAY & ORAM
```

```

; STREET: 655 Fifteenth Street, N. W., Suite 330 - G
; STREET: Street Lobby
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,218
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: KITTS, Monica C.
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P1615-5002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-396-218-2
```

```

Query Match          7.2%; Score 156; DB 1; Length 422;
Best Local Similarity 23.8%; Pred. No. 6.1e-08;
Matches 97; Conservative 51; Mismatches 202; Indels 58; Gaps 15;
```

```

QY 39 QEAMAVLQESNVDLVWTRCNGHMIATRGQLIREAVEDYRHSSECPFIPEGAEAYDF 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 31 REAGPVV-EVNAP-----AGPRAMVITDDALAREVLADRFVKD-----PDLAPAMRG 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 99 IPTSMDE--PREQOFRLA-----NQVGVAPVD-----KLENRIQELACSLIESL- 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 79 VDGLDIVEPELRPTLLIADGEAHRRRLRIHAPAFNPRLAERTDRIAAIAGRLTELA 138
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 143 ----RPGQCNETEDYAEPPRIPIFMLAGLP-----EEDIPHLKYL-----TDQ 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 139 DASGRSGKRAELIGGFAYHPRFLVLCGLGVPTDPMAREAVSVLKALGLGPGSGGD 198
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 MTPRDSMTFAEAKALYDYLPIIEQRQKPGTDAISIVANGQVNGRPTSDEAKRMCG 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 199 GTDPAGVPRPTSLES--LLEAVHSARRDPTMTVLYERAQAEFGVSDDQLVVMIT 256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 LLVGLDITVYNFLFSMEFLAKSEHROELIERPERIPACEELLRFSLV--ADGRL 301
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 257 GLIFAGHDTTGSFLGF--LLEAVLAGRLAADADDAVSRFYEELRHHRPVYTLWRFPA 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 302 TSDYEHFGVOLKKGDQILLPQMLSGLDERENACPMHVDPSHOKVSHTTFGHSHLCLGHL 361
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 314 ATEVTTIGVLRPRKAPVLDIEGTDTGRIHNDAPHAHFPRDRSRRLTFGDDPHYCTIGEQ 373
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 362 IARREIIVTLKEMLTRIPDFSIA--PGAQIQ--HKSGLVSGVALPLVW 406
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 374 LAQLESRTMIGVLSRFPQARLAVPYDELRLWCRKGAQTARLTDLF-VW 420
```

```

RESULT 14
US-08-760-116-2
; Sequence 2, Application US/08760116
; Patent No. 5786190
; GENERAL INFORMATION:
; APPLICANT: INVENTI, Augusto
; APPLICANT: BREME, Umberto
; APPLICANT: COLOMBO, Anna L.
; APPLICANT: HUTCHINSON, Charles R
```

```

APPLICANT: OTTEN, Sharee
APPLICANT: SCOTTL, Claudio
TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIKAIKO, MARCELSTEIN, MURRAY & ORAM LLP
STREET: 655 Fifteenth Street, N. W., Suite 330 - G
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/760,116
FILING DATE: 3-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/396,218
FILING DATE: 27-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: KITT'S, Monica C
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: P1615-6007
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/638-5000
FAX: 202/638-4810
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-760-116-2

```

Query Match 7.2%; Score 156; DB 1; Length 422;  
Best Local Similarity 23.8%; Pred. No. 6.1e-08;  
Matches 97; Conservative 51; Mismatches 202; Indels 58; Gaps 15;

```

39 OEAMAVLQESNPDLVWTRNGMHWATRGQILREAYEDYRHFSSECPPIPRAGEAYDF 98
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
31 REAGPYV-EVNAF-----AGGPWVITDDALAREVLADPRFVKD-----PDLAAPAMRG 78
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
99 IPTSMD--PPEORQFRALA-----NOVGMPPVD-----KLENRIQELACSLIESL- 142
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
79 VDDGLDIPVPELRFPLLIADVGEAHRRLRIHAFNPRLAERTDRIAAIAGRLITELA 138
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
143 ----RPGQGNFTEDYAEFPRIEFLMLAGLP-----EEDIPHLKYV-----TQ 183
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
139 DASGRGKPAELIGFRAYHPILVICELGVPTDPMAREAVSVKALGLGPGSGGD 198
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
184 MTRPDSMTFAEAKALDYLLPIIEOROKPGTDAISIVANGVNGRPITSDAKRMCG 243
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
199 GTPDAGVDTSLTESL--LLEAVHSARNDPTMTRVLYERQAQFGSVSDQLVYMIT 256
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
244 LLVGLDTVNVFLFSMEFLAKSPHROELIERPERIPACHELLRRLSLV--ADGRIL 301
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
257 GLIFGHDYTGSLGF---LLAEVLAGRLAADADAVSRFVEALRYHPVPYVYTLMPRA 313
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
302 TSDYERHGOVLKKGDQILLPQMLSGDERENACPMHVDSSROKVSITTGSHGLDQGH 361
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
314 AEEVITGVRLPRGAVLVLDICTNTDGRHNDAPPAFHPDRPSWRRLTGGDGHYIGEQ 373
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
362 LARREITVLKEMLRIPDFSLA--PGAQIQ--HKSGIVSGVALPLWV 406
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
374 LAQLESRMIGVLRSPFARLAVPDELRLMCRKKGQOTARITELP-VV 420
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

```

```

RESULT 15
US-08-194-981E-5
Sequence 5, Application us/08194981E
Patent No. 5886157
GENERAL INFORMATION:
APPLICANT: GUENGERICH, F. Peter
APPLICANT: GUO, Zuyu
APPLICANT: SANDHU, Punam
APPLICANT: GILLAM, Elizabeth M. J.
TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF
TITLE OF INVENTION: HUMAN
TITLE OF INVENTION: CYTOCHROME P450
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: Suite 1200, 127 Peachtree Street, NE
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303-1811
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194,981E
FILING DATE: February 10, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Elizabeth Selby
REGISTRATION NUMBER: 38,298
REFERENCE/DOCKET NUMBER: 22000.0022
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 688-0770
FAX: (404) 688-9880
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 512 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: N-terminal
US-08-194-981E-5

```

Query Match 5.9%; Score 129; DB 2; Length 512;  
Best Local Similarity 22.4%; Pred. No. 5.4e-05;  
Matches 95; Conservative 67; Mismatches 158; Indels 104; Gaps 22;

```

27 DWVNPNSLQAGVQEMAVLQESNPDLVWTRNGMHWATRGQILREAYEDYRHFSSECPPI 81
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
107 DLYTFPLISNGQSMSPS-----PD-----SGPVMAAR-----RLAONGKLSFSIASD 149
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
82 ----SSSECPPIPRAGEAYDFIPTSD--PPEORQFR-----ALANQVGMPPVDKLENR 130
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
150 PASSTCYIEIENHVSRAEVLISITQELMAGPQHFNRYRVVSVYNVICAIGRGRIYDIN 209
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
131 IDELACSLIESLRPGQGNFTEDYAEFPRIEFLMLAGLPEDDIPHLKYVLTQDMTRPDGS 190
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
210 HQEL-LSLV-----NLNNNGE-----VVGSGNPAEIPILRYL-----PNPS 246
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
191 M-TFAAKALDYLLPIIEORQ-----KPGDPAISIVANGQ-----VNGRPTSD 237
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
247 LNAFKDLNKFYSFMQKVKVKEHYKTFEKGHINDYD--SLIEHCEKQOLDENAVQLSDE 304
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
238 AKRMGLLVG--GLDVNVNLFSPMEFLAKSPHROELIE-----RP-----ERI 281
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
305 KIINIYVLDLFGAGFDVTVAISNLMYLVNMPVOKIOEEDLDYIGRRRRLSDRSHL 364
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
282 P---AACCELRLRRLSLV--ADGRILTSDEYFHGOVLKKGDQILLPQMLSGDERENACP- 335
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

```

Db 365 PYMEAFIETFRHSSEVPFTIPHSTRTDTSIKGFYIPKRCVFVNQMOINHDOKLWVNS 424  
QY 336 -----MHVDFSROKY---SHTTGHSHLCLGQHLARREIIVTLKKEWLTTRIPDFSIAP 385  
Db 425 EFLPERELTPDGAIDKVLSEKVIIFGMRKRCIGETVARMEVFLFALAILLQV-EFSVPL 483  
QY 386 GAQI 389  
Db 484 GVKV 487

Search completed: October 4, 2000, 12:38:03  
Job time: 21420 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 05:49:35 ; Search time 73.09 Seconds  
(without alignments)  
350.531 Million cell updates/sec

Title: US-09-246-451-2  
Perfect score: 2180  
Sequence: 1 TTERIQSNANLAPLEPHVPE.....IVSGVQALPLVWDPATTKAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_64:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2180	100.0	415	1	O4PSCP
2	380	17.4	393	1	C70829
3	371	17.0	410	2	S39924
4	365	16.7	397	1	S51594
5	362	16.6	396	1	B69851
6	360	16.5	398	1	H70807
7	346.5	15.9	404	2	T30231
8	340	15.6	403	2	B35401
9	334.5	15.3	410	1	O4B56M
10	334.5	15.3	428	1	F70729
11	333	15.3	410	1	E69611
12	331.5	15.2	405	1	B42606
13	330	15.1	406	2	A35401
14	329	15.1	411	1	JC5859
15	326	15.0	406	1	A48495
16	324.5	14.9	376	1	G69679
17	320.5	14.7	405	1	H70752
18	318.5	14.6	395	1	G69594
19	318	14.6	310	2	T44767
20	318	14.6	428	1	A42971
21	316.5	14.5	399	1	A53578
22	316	14.5	412	2	S24750
23	313.5	14.4	410	2	JC4287
24	304	13.9	417	2	S49051
25	304	13.9	438	1	E70515
26	303	13.9	402	2	JC5151
27	303	13.9	404	1	JC5150
28	296.5	13.6	408	2	PD0007
29	296.5	13.6	408	2	JC5674

30	294.5	13.5	406	1	S18531	cytochrome P450 er
31	293.5	13.5	337	2	T30601	cytochrome P450 hy
32	291.5	13.4	433	1	B70677	cytochrome P450 rv
33	285	13.1	381	1	S15809	cytochrome P450 cy
34	276.5	12.7	310	2	T44857	probable hydroxyla
35	266	12.2	412	1	B40634	erythromycin monoo
36	265.5	12.2	411	2	T36526	probable cytochrom
37	265	12.2	414	1	E70708	cytochrome P450 rv
38	261.5	12.0	386	2	T30235	cytochrome P450 -
39	254	11.7	401	1	I40208	cytochrome P450 BJ
40	254	11.7	406	3	T17487	cytochrome P450 hy
41	248	11.4	411	1	JC4403	cytochrome P450 -
42	240	11.0	433	2	T44587	cytochrome P450 ho
43	236	10.8	402	1	A70707	cytochrome P450 rv
44	235	10.8	396	1	H70730	cytochrome P450 rv
45	225	10.3	391	3	T17486	cytochrome P450 hy

## ALIGNMENTS

RESULT 1

O4PSCP  
camphor 5-monooxygenase (EC 1.14.15.1) cytochrome P450 101 - Pseudomonas putida plasm

N:Alternate names: cytochrome P450-CAM

C:Species: Pseudomonas putida

C:Date: 30-Apr-1982 #sequence.revision 31-Dec-1993 #text\_change 03-Mar-2000

C:Accession: A25660; S34614; C60886; A00194

R:Unger, B.P.; Gunsalus, I.C.; Sligar, S.G.

J. Biol. Chem. 261, 1158-1163, 1986

A:Title: Nucleotide sequence of the Pseudomonas putida cytochrome P-450-cam gene and

A:Reference number: A94678; MID:86111751

A:Accession: A25660

A:Molecule type: DNA

A:Residues: 1-415 <UNG>

A:Cross-references: GB:M12546; MID:g151114; PIDN:AAA25760.1; PID:g151115

R:Atamaki, H.; Koga, H.; Sagara, Y.; Hosoi, M.; Horiochi, T.

Biochim. Biophys. Acta 1174, 91-94, 1993

A:Title: Complete nucleotide sequence of the 5-exo-hydroxycamphor dehydrogenase gene

A:Reference number: S34613; MID:93326643

A:Accession: S34614

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-42 <ARA>

A:Experimental source: ppg1, ATCC 17453; CAM plasmid

R:Romeo, C.; Moriwaki, N.; Yasunobu, K.T.; Gunsalus, I.C.; Koga, H.

J. Protein Chem. 6, 253-261, 1987

A:Title: Identification of the coding region for the putidaredoxin reductase gene fro

A:Reference number: A60886

A:Accession: C60886

A:Molecule type: DNA

A:Residues: 408-415 <ROM>

R:Hanley, M.; Ames, L.G.; Yasunobu, K.T.; Shastri, B.A.; Gunsalus, I.C.

J. Biol. Chem. 257, 42654-42671, 1982

A:Title: Amino acid sequence of the Pseudomonas putida cytochrome P-450. II. Cyanogen

A:Reference number: A00194; MID:83030788

A:Accession: A00194

A:Molecule type: protein

A:Residues: 2-55, 58-276, 'Q', 278-361, 'S', 363-407, 'N', 409-415 <HAN>

C:Genetics:

A:Gene: camC; CYP101

A:Genome: Plasmid

C:Function:

A:Description: catalyzes hydroxylation of camphor to yield 5-exo-hydroxycamphor; elec

C:Superfamily: Pseudomonas plasmid camphor 5-monooxygenase; cytochrome P450 homology

C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxigena

F:246-380/Domain: cytochrome P450 homology <CYP>

F:358/Binding site: heme iron (Cys) (axial ligand) #status experimental

Query Match 100.0%; Score 2180; DB 1; Length 415;  
Best local Similarity 100.0%; Pred. No. 7, 6e-159;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	TTETIQSANANLAPLPHPNPHPLVDFEDFMYNPSNLSAGVQZQEMAVYLQESNVDDLWYTRCNG	60
Db	2	TTETIQSANANLAPLPHPNPHPLVDFEDFMYNPSNLSAGVQZQEMAVYLQESNVDDLWYTRCNG	61
Qy	61	GHWIATROQLREAEADYRHHSSSECPFIPRAGEAYDPIPSMOPPEORORALANOVYG	120
Db	62	GHWIATROQLREAEADYRHHSSSECPFIPRAGEAYDPIPSMOPPEORORALANOVYG	121
Qy	121	MPVVDKLENRIQELACSLIESLRPOGQCNFTEDYAERPPIRIFMLAGLPREDIPHLKYL	180
Db	122	MPVVDKLENRIQELACSLIESLRPOGQCNFTEDYAERPPIRIFMLAGLPREDIPHLKYL	181
Qy	181	TDQMRPRGQSMTFARAKALDYLLPIITEOROKRGTDATISIVANGOVNGRPITSDEAKR	240
Db	182	TDQMRPRGQSMTFARAKALDYLLPIITEOROKRGTDATISIVANGOVNGRPITSDEAKR	241
Qy	241	MGCLLVAGGLTVVNFLEFSMEFLAKSPENHOELIERPERIPRACEELLRRFSLVADGRI	300
Db	242	MGCLLVAGGLTVVNFLEFSMEFLAKSPENHOELIERPERIPRACEELLRRFSLVADGRI	301
Qy	301	LTSDFEFGVOLKKGDOILLPQMLSGLDERENACPMHVDFSROKUVSHTEGHSGLCLGQ	360
Db	302	LTSDFEFGVOLKKGDOILLPQMLSGLDERENACPMHVDFSROKUVSHTEGHSGLCLGQ	361
Qy	361	HLAREIIVTLKEMLTRIPDPSIABGAOIOHKSGIVSCVOLALPYMDPATTKAV	414
Db	362	HLAREIIVTLKEMLTRIPDPSIABGAOIOHKSGIVSCVOLALPYMDPATTKAV	415

RESULT 2  
C70929  
cytochrome P450 RV1785c - Mycobacterium tuberculosis (strain H37Rv)  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Mycobacterium tuberculosis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: C70929  
R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: C70929  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-393 <COL>  
A:Cross-references: GB:A022021; GB:AL123456; NID:3250699; PIDN:CAAI7707.1; PID:e125459  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV1785c  
;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
;Keywords: oxidoreductase

Query Match	17.4%;	Score 380;	DB 1;	Length 393;
Best Local Similarity	28.7%;	Pred. No. 1.5e-21;		
Matches 110;	Conservative 71;	Mismatches 182;	Indels 20;	Gaps 10

OY		31	PSNLSAQGEAMVLOESBNVDLWYTRONGHWTATROQLIREAVEDYRHHSSCEPIFR	90
		:	:   :	:
		:	:   :	:
Db		23	PMAADRGV--GWKTLRRDAG--PVVF--MNGWYYL-TREDYLAALRNPKVSSSKALOP-	74
OY		91	EAGEAYDFIPTSMPEEROFERFALANQVGMPVYDKLENRQELACSLIESLRQGOCNF	150
		:	:   :	:
		:	:   :	:
Db		75	-PGNPPLPVIAPDDPEETRYKRIILOPYFSPAALKALPSIKRRHTVAVIDIAGRGCEA	133
		:	:   :	:
		:	:   :	:
OY		151	TEDYAEPRIIFIMLLGLPEDIPHLKYLTDOM----TRDGSGMTFAEAKEALYDLIP	206
		:	:   :	:
		:	:   :	:
Db		134	MADLANLEFPOLFLVLYGLPRBEDRDRLIGMKWDYIAMSDRHPHEADYAAARELLETILTA	193
OY		207	IIEORROKPGTDAISIVANGGVNGRPITTSDEAKRMCGILLVGLDITYVNFLSFMELAK	266

```

Db      194  MVARRRNPNPGDVLGYQ---QIGEDPLSELEVLGSHLLILAGLDYVRAVAFSLLELAR 250
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      267  SPEHRELIERPERETIPACCELLR-RFSLVADGRILITSPYEFHGVLQKKGOILLPOMLS 325
      251  RPOLRAMLRNPKQIRFEIIEILRLEPSAVAPRVTEVTVGCMTLPAGSPVRLCMAMV 310
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      326  GLDERENACPMHVDSPKQVSHHTPCGHSLCIGOLHAREIITVLKLEWILPIPFSTAP 385
      311  NRDSDDAMSTDELVMQCKVIRHMGFGGGRILRCLGSHLARELTLLVGEWLNQIPFELAP 370
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      386  --GAQIQHKSGLVSGVQALPLVM 406
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      371  DYAEELRFPKSFSA-LKNLPLRW 392
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 3  
S39924  
cytochrome P450meg - *Bacillus megaterium* (ATCC 13368)  
C:Species: *Bacillus megaterium*  
A:Variety: ATCC 13368  
C:Date: 07-Oct-1994 #sequence\_revision 23-Feb-1996 #texl\_change 04-Mar-2000  
C:Accession: S39924; S32216  
R:Rauschenbach, R.; Isernhagen, M.; Noeske-Jungblut, C.; Boldol, W.; Siewert, G.  
Mol. Gen. Genet. 241, 170-176, 1993  
A:Title: Cloning sequencing and expression of the gene for cytochrome P450meg, the st  
A:Reference number: S39924; MUID:94049677  
A:Accession: S39924  
A:Molecule type: DNA  
A:Residues: 1-410 <RAU>  
A:Cross-references: EMBL:Z21972; NID:g2882296; PTDN:CAA79985.1; PID:g288300  
A:Experimental source: ATCC 13368  
C:Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; heme; iron; metalloprotein  
C:355/Binding site: heme iron (Cys) (axial ligand)#status predicted

Query Match	17.0%;	Score 371;	DB 2;	Length 410;
Best Local Similarity	30.1%;	Pred. No. 7.9e-21;		
Matches 109;	Conservative 53;	Mismatches 180;	Indels 20;	Gaps 7;

QY	63	WIATRGOLIRATVADYDYNHFSSE-----CCPIREAGEAYDFIP-TSMDDPEOROPRALA	115
Db	44	WNVREYEDYVKVVLSDYHNFHFFSVKRRITTSVGTDSSEGSVEPKIQTIESDDPHRRKRSLL	103
QY	116	NOVGMVPVVDKLENRIQELACSLTESLPQOCNFTEADVAEPPIRIFMLLAGIPREDIP	175
Db	104	AAATPSRLQWMERIRIDEINDELIGQNDGSGTEIDIVASLASPLRTIYMADLMKVPKSDRL	163
QY	176	HLKYLTPDMTRPDGSMTPAE-----AKELADYLLPIIEQRKRGPTDAISIVANGQV	228
Db	164	LEKKWVDLLELPREDREKOEYDKLQVAAREYUQYELPIYQKRRLPADDIISDLLKSEV	223
QY	229	NGRPTTSLEAKRMGGLLVGSLDTPVNVFSLFSM-EFLAKSPENROELIERPEKIPAACEE	287
Db	224	DGEMPTDDEVVRTMILLGAGVETTSHLANSFYSLLDYDKUYVQDELHEMLDIPQAVEE	283
QY	288	LIR-RESLVADGRITSDYEFHNGVQLKKGOILLPQMLSGIDEREENACPMYHDFSR-QKV	345
Db	284	MLRRFENLIKIDRIKVEDNDNLGSELKEGSSVYVVMASAMNDMEFEDPPTLIINHPRNK	343
QY	346	SHHTFGHSHLCLQHLARRELIYTLKEWLTIRIDFSIABGAOIOHK--SGIVSGVOAL	402
Db	344	KHLTFGNRPHTCSGLAPLARLEAKIALALAPLKKFKHIEAVPSQLEMLTDSATGQTLTSL	403
QY	403	PL 404	
Db	404	PL 405	

cytochrome P450 mycG - Micromonospora griseorubida

N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Micromonospora griseorubida  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S51594  
R:Inouye, M.; Takada, Y.; Muto, N.; Beppu, T.; Horinouchi, S.  
M01. Gen. Genet. 245, 456-464, 1994  
A:Title: Characterization and expression of a P-450-like mycinamicin biosynthesis gene u  
A:Reference number: S51593; MUID:95107242  
A:Accession: S51594  
A:Molecule type: DNA  
A:Residues: 1-397 <INNO>  
C:Cross-references: EMBL:D16098; NID:g286050; PIDN:BA03672.1; PID:d1004187; PID:g303644  
C:Genetics:  
A:Gene: mycG  
A:Start codon: GTG  
C:Superfamily: Bacillus cytochrome P450 CYP106, cytochrome P450 homology  
C:Keywords: heme; oxidoreductase  
C:231-368/Domain: cytochrome P450 homology <CYP>

Query Match	16.7%;	Score 365;	DB 1;	Length 397;
Best Local Similarity	28.5%;	Pred. No. 2.2e-20;		
Matches 105;	Conservative 57;	Mismatches 168;	Indels 38;	Gaps 8;

```

QY      63 WIAIRGQLIREAYEDYRHFSECFI-----PRAGEAYDFIPTSMDPPEQGRFA 113
      1: 11 111 1 : : 11 1 11111 : 1
Db      44 WLVR-----YEDYRAVLGDRFVAGCPMTREDFRPRPEMVGGLSMDPPEHSRLR 96

```

```
QY      114 LANVVGMPVVDKLENRIGQLACSLIESLRPGQ-CNFTEDYAEPPIRIIFMLACLP EE 172
         | : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      97 LVYKAFTARAESLRRPARELAHEDIOMAAAGOPADLVAMFAROLPVPRICELLVGPSA 156
```

```
QY 173 DIPHLKYLTDQMTRPDGS-----TFAEAKALDYLIPIEQROKPGTDAISV 223
      | : | | : | | | : : | | : | : |
Db 157 D-----HRTFRSGAFLSTAETVAEEMOEAEOAAYMGDLDRRRREPTDVSAL 209
```

QY 224 ANGVNGRPITSDEAKRMCGLLVGGDTVNFLSFSMEFLAKSPENRQELIERPERIPA 283  
:  
DB 210 VOAROODSTSEOLFIDLALIGILVAGYSTMOIADFVILMTREIRRLRIDRPDELPS 269  
:

QY 284 ACEELLRRPSL---VADGRILTSDYERHGVOLKKDQIILLPOMLSGLDERENACPMHWDF 340  
| | | | | | : : : : :  
Db 270 AVEFETRWVDELGVCEAFBPYAVEDVTIGVETPAGEPVASTGAANPDQAOPENDADITDV 329

OY 341 SROKVSHTTPEHGSHLCLGOMHARELITVTLKEMLTRIPDFSIA-PCGAQIQKSG-IVSG 398  
| | |||| | || | : | : | : | : | : | : | : |  
Db 330 ДПРОМОНТ ССЕСУВНУСТСАД АДУЕЛОВАЛ ЕМУЛОПДОБРАСТЕНМОА БАСЕСУТИ БО 300

QY	399	VQALPLVM	406
		:	
DB	300	PLPLPLVM	307

## RESULT 5

Cytochrome P450 yj1B - *Bacillus subtilis*  
N:contains: oxidoreductase (EC 1.-.-.-)  
C:Species: *Bacillus subtilis*

C:Accession: B69851  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Aveve, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Car-

Nature 390, 249-256, 1997  
 Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galiz  
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.

A: Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masu  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.  
Rieger, M.; Riolto, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata,

A:Authors: Yoshikawa, H.F.; Zunshtein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis* strain 168  
A:Reference number: A69580; MUID:98044033  
A:Accession: B69851  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-396 <KUN>  
A:Cross-references: GB:J29110; GB:JL009126; NID:g2633472; PIDN:CAB13078.1; PID:el18322  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yjiB  
C:Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase  
E:236-371/Domain: cytochrome P450 homology <CYP>

Query Match	16.6%;	Score 362;	DB 1;	Length 396;
Best Local Similarity	30.1%;	Pred. No. 3.6e-20;		
Matches 99;	Conservative 55;	Mismatches 157;	Indels 18;	Gaps 7;

[illegible]

QY 131 IQELACSLIESLRPGQCNFTEDYAEPPFIRIFMLAGLPEEDIPHLKYLTDQT-----1859  
 111: 111: : : : 111: 111: : : :  
 Db 112 IQEITDELIOFGGRSEFDLPHFYSPLPEVIVISRLGVSAMHEQFAWSLVLSTPKD 1711

QY 186 -RPDGSMTFAE----AKEALDYDLPIITFEOROKPPTDAISIVANGOVNCRPTSDEAKR 240  
 : | | : | | : ||| : | | | | : : : :  
Db 172 KSELEKAFLEERDKCEEELAAEFAGLIEEKRNKPEDIISTIVEAEFTGEXSGFELIP 231

QY 241 MCGLLVGGLDTVNFLSFSMEFLAKSPHROELIERPERIPACCELLR-RFSLVADGR 2999  
| ||||| : : : : : || : : : : || || || |  
Db 232 ECTLLLVAGNEFTNTLTSNAMYSTFTPEVYFRLRSHPELMPDVAEEALREAPAPVILRR 2911

Dh 300 ILTSDYEHGVOYLKKQDQILLPQLSGIDERENACPMHNVFSRQKVSHTTTEGHSHCLG 359

292 TAKPRDETGHLIKFEGMVLAFVASANPREAKFDPRMFQIYBRNDHTAEFGHTHEIC 354

QY 360 QHAREIYVLEKWLTRIPDF---SIAP 385  
 ||| | : | | |  
 Dn 352 ADIAPREANTAIWSTSAEDYHVEGUSTIB 380

## RESULT 6

cytochrome P450 RV3518C - Mycobacterium tuberculosis (strain H37RV)  
N:Contains: oxidoreductase (EC 1.-.-.)  
C:Species: Mycobacterium tuberculosis

C;Accession: H70807  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnter, T.; Churcher, C.; Harris,  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N

Nature 393, 537-544, 1998  
A; Authors: Sqaates, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, A.; Title: Deciphering the biology of *Mycobacterium tuberculosis* from the cc

A:Accession: H70807  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CA17755.  
A:Experimental source: strain H37Rv  
C:Genetics:

C:Superfamily: *Bacillus cytochrome P450 CYP106*; cytochrome P450 homology  
C:keywords: oxidoreductase

Query Match	16.58;	Score 360;	DB 1;	Length 398;
Best Local Similarity	28.28;	Pred. No. 5.2e-20;		





A: Molecule type: DNA  
A: Residues: 1-410 <BEL>  
A: Cross-references: EMBL:Y11043; PIDN:CA71937.1  
A: Experimental source: strain 1A1  
C: Genetics:  
C: Gene: cypA  
C: Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C: Keywords: oxidoreductase  
F: 245-381/domain: cytochrome P450 homology <CYP>

Query Match	15.3%	Score	333;	DB	1;	Length	410;
Best Local Similarity	27.2%;	Pred. No.	6.2e-18;				
Matches	86;	Conservative	69;	Mismatches	139;	Indels	22;
				Gaps			5.

[illegible]

```

Query Match 15.2% Score 331.5: DB 1, Length 405;
Best Local Similarity 25.6%: Pred. No. 8e-18;
Matches 107; Conservative 73; Mismatches 165; Indels 53; Gaps 10;

18 VPEHLEVD---FDMYNPSNLSAGVOEAMAAVLQESNVPDLVWTRCNG-GHWIATRCQLIR 72
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6 VPDLLAFQADDAFAQODHN-----RARRRREPVPORI--RTNGDGLDAWLTITREDDYK 53
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY

```

```
QY      73 EAYEDYNHSECECFIPREAGEAVIDPISM-----DPEQRFR    112
       :| | | | | | | | | | | | | | | | | | | | | | | |
Db      54 QALLD-----PRIAKDRGTOIIIEKRKLADERRRPFSDLCRHMLNTDPDHRLR   105
                                     :| | | | | | | |
QY     113 ALANQVGMPEVDKLENRIDELACLSELRPGQCNFREDYAEPPIRFIMLAGLPBE   172
               :| | | | | | | | | | | | | | | | | | | | |
Db     106 KLYVKATFARVEBLRRRIEQITDDLLDLRLAGRSEVDLIDEAFPLPTIVISELMGVDS   165
                                               :| | | | | | |
QY     173 DIPHLKYLTDMTRPDGSMTFAAKE--ALDYULPIIBOROKFGTDAISIVANGQVNG   230
               :| | | | | | | | | | | | | | | | | | | | |
Db     166 RRDFRSMTNVLV--DOSGEQAQASVANAVEVELTELIAKKRTBEGDDLTLALLEAVBDG   223
                               :| | | | | | | | | | | |
QY     231 RPITSDAKKRMCGILLVGGLDIVYNFLSTSMFEFLAKSPHRQELLIEREPRIPAACEELLR   290
               ::| | | | | | | | | | | | | | | | | | | | |
Db     224 DRLESGETILAMVELLIWAGHEHTTVNLIGNCVSLIGNSPDQLAALRNDESSLPGAILEETLR   283
                       :| | | | | | | | | | | | | | | |
QY     291 RFSLYAOG--RILTSDYEFIGVOLKKGDDOLLPMLSGLDERENAKCPMHNDFSROKYSHT   348
               ||| | | | | | | | | | | | | | | | | |
Db     284 YESYVANGTEFHRAEAVERFEDDVYIPGSELVMAVALGANNDGERFEDPPRFDTIRETTGHV   343
                                   :| | | | | | | |
QY     349 TTFGGSHSGICGOHLAREELLVLTKEMULTIRPOESTA--PGAIQIKKSIGSVGYOALPL   404
               ++++++||| | | | | | | | | | | | | | | |
Db     344 AFGGIHFVCYGAAALARLEAOIYAGGRLLERPPDRMNASPDULRWRFVSVMRGLEKLPV   401
```

[illegible]







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:36:49 ; Search time 27.02 Seconds  
(Without alignments)  
474.815 Million cell updates/sec

Title: US-09-246-451-2

Perfect score: 2180  
Sequence: 1 TTTETIQSNANLAPLPHPVE.....IVSGVOALPLVWDPATTKAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2180	100.0	414	1 CPXA_PSEPU	P00183 pseudomonas
2	595.5	27.3	414	1 Y4VG_RHISN	053215 rhizobium s
3	380	17.4	393	1 YH85_MYCTU	053936 mycobacteri
4	371	17.0	410	1 CPXM_BACME	006069 bacillus me
5	362	16.6	396	1 YJTB_BACSU	034374 bacillus su
6	360	16.5	398	1 YZ18_MYCTU	053563 mycobacteri
7	357.5	16.4	405	1 CPXM_BACSU	P27632 bacillus su
8	340	15.6	402	1 CPXF_STRGO	P18327 streptomyc
9	334.5	15.3	410	1 YM66_MYCTU	050696 mycobacteri
10	334.5	15.3	428	1 CPXY_BACSU	008469 bacillus su
11	333	15.3	410	1 CPXK_SACER	P33271 saccharopol
12	331.5	15.2	405	1 CPXK_SACER	P18327 streptomyc
13	330	15.1	405	1 CPXK_SACER	P18327 streptomyc
14	320.5	14.7	405	1 YC56_MYCTU	P53554 bacillus su
15	318.5	14.6	395	1 BIOT_BACSU	P33006 pseudomonas
16	318	14.6	428	1 CPXL_PSEPU	P23295 fusarium ox
17	316.5	14.5	399	1 FAST_RHOFA	P46373 rhodococcus
18	316	14.5	412	1 CPXH_STRGO	P26911 streptomyc
19	304	13.9	438	1 Y180_MYCTU	008464 mycobacteri
20	303	13.9	402	1 NOR_FUSOX	P23295 fusarium ox
21	296.5	13.6	408	1 NOR_CYLTO	012599 cyllindroc
22	293	13.4	381	1 CPXG_STRGO	P23296 streptomyc
23	291.5	13.4	433	1 YZ45_MYCTU	P71856 mycobacteri
24	289.5	13.3	403	1 CPXJ_SACER	000441 saccharopol
25	288.5	13.2	403	1 NOR_CYLTO	000465 cyllindroc
26	277.5	12.7	397	1 CPXO_SACER	P48635 saccharopol
27	266	12.2	400	1 CPXP_RHISN	P55544 rhizobium s
28	265	12.2	414	1 P778_MYCTU	P77993 mycobacteri
29	254	11.7	401	1 CPXP_BRAJA	059203 bradyrhizob
30	236	10.8	402	1 YM76_MYCTU	P77992 mycobacteri
31	235	10.8	396	1 YH76_MYCTU	059571 mycobacteri
32	224	10.3	436	1 THCB_RHOER	P43492 rhodococcus
33	221.5	10.2	400	1 YV21_MYCTU	008362 mycobacteri

34	218	10.0	422	1 CPXC_AGR6	P24466 agrobacteri
35	215	9.9	429	1 CPXR_BRAJA	059204 bradyrhizob
36	207.5	9.5	489	1 CPXB_MYCTU	059572 mycobacteri
37	195	8.9	405	1 CPXP_BACSU	034926 bacillus su
38	189	8.7	467	1 CPXA_RHISN	P55543 rhizobium s
39	181	8.3	368	1 CPXM_SULSO	055080 sulfolobus
40	162	7.4	1048	1 CPXB_BACME	P14779 bacillus me
41	156.5	7.2	503	1 CP30_SHEEP	029496 ovis aries
42	153.5	7.0	531	1 CP27_HUMAN	002318 h cytochrom
43	152	7.0	492	1 CP53_PIG	002390 sus scrofa
44	151	6.9	407	1 CPXD_AGR6	P24467 agrobacteri
45	149.5	6.9	489	1 CP33_RABIT	P00182 oryctolagus

## ALIGNMENTS

RESULT	1	STANDARD	PRT	414 AA.
CPXA_PSEPU				
ID	CPXA_PSEPU			
AC	P00183:			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	CYTCHROME P450-CAM (EC 1.14.15.1) (CAMPOR 5-MONOOXYGENASE).			
GN	CAMC OR CYP101.			
OS	Pseudomonas putida.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;			
OC	Pseudomonas.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 8611751.			
RA	Unger B.P., Gunsalus I.C., Sligar S.G.:			
RT	"Nucleotide sequence of the Pseudomonas putida cytochrome P-450cam			
RT	gene and its expression in Escherichia coli."			
RL	J. Biol. Chem. 261:1158-1163(1986).			
RN	[2]			
RP	SEQUENCE OF 385-414 FROM N.A.			
RC	SPRAIN-ATCC 17453:			
RX	MEDLINE: 90130389.			
RA	Koga H., Yamaguchi E., Matsunaga K., Aramaki H., Horiuchi T.:			
RT	"Cloning and nucleotide sequences of NADH-pyridine oxidoreductase			
RT	gene (camA) and putidaredoxin gene (camB) involved in cytochrome			
RT	P-450cam hydroxylase of Pseudomonas putida."			
RL	J. Biochem. 106:831-836(1989).			
RN	[3]			
RP	SEQUENCE:			
RX	MEDLINE: 83030788.			
RA	Hanlin M., Ames L.G., Yasunobu K.T., Shastri B.A., Gunsalus I.C.:			
RT	"Amino acid sequence of the Pseudomonas putida cytochrome P-450. II.			
RT	Cyanogen bromide peptides, acid cleavage peptides, and the complete			
RT	sequence."			
RL	J. Biol. Chem. 257:12664-12671(1982).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).			
RX	MEDLINE: 86143817.			
RA	Fitzel B.C., Weber P.C., Hardman K.D., Saleme F.R.:			
RT	"Structure of ferricytochrome c' from Rhodospirillum rubrum at			
RT	1.67-A resolution."			
RL	J. Mol. Biol. 186:627-643(1985).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).			
RX	MEDLINE: 86059514.			
RA	Poulos T.L., Fitzel B.C., Gunsalus I.C., Wagner G.C., Kraut J.:			
RT	"The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450."			
RL	J. Biol. Chem. 260:16122-16130(1985).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).			
RX	MEDLINE: 98019009.			
RA	Schlichting I., Jung C., Schulze H.:			
RT	"Crystal structure of cytochrome P-450cam complexed with the (1S)-			
RT	camphor enantiomer."			
RL	FEBS Lett. 415:253-257(1997).			

```

RN      [7]
RP      X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX      MEDLINE; 98313255.
RA      Vlakovic M., Sligar S.G., Li H., Poulos T.L.;
RT      "Understanding the role of the essential Asp251 in cytochrome p450cam
RT      using site-directed mutagenesis, crystallography, and kinetic solvent
RL      isotope effect.";
RN      Biochemistry 37:9211-9219(1998).
[8]
RP      STRUCTURE BY NMR.
RX      MEDLINE; 97459726.
RA      Mourou C., Bondon A., Symoneaux G., Jung C.;
RT      "H-NMR study of diamagnetic cytochrome p450cam: assignment of heme
RT      resonances and substrate dependence of one cysteinate beta proton.";
RL      Febs Lett. 414:203-208(1997).
CC      -I- FUNCTION: INVOLVED IN A CAMPHOR OXIDATION SYSTEM.
CC      -I- CATALYTIC ACTIVITY: CAMPHOR + PUTIDAREDOXIN + O(2) = 5-EXO-
CC      HYDROXYCAMPHOR + OXIDIZED PUTIDAREDOXIN + H(2O).
CC      -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; M12546; AAA25760.1; -.
DR      EMBL; D000528; BAA00412.1; -.
DR      PIR; A23660; O4PSCP.
DR      PDB; 2CXP; 15-APR-91.
DR      PDB; 3CPP; 15-APR-91.
DR      PDB; 4CP4; 15-JUL-91.
DR      PDB; 5CPP; 15-JUL-91.
DR      PDB; 6CPR; 15-JUL-91.
DR      PDB; 7CJP; 15-JUL-91.
DR      PDB; 8CJP; 15-JUL-91.
DR      PDB; 1CP4; 15-JAN-93.
DR      PDB; 2CP4; 15-JAN-93.
DR      PDB; 3CP4; 15-JAN-93.
DR      PDB; 4CP4; 15-JAN-93.
DR      PDB; 5CP4; 16-SEP-98.
DR      PDB; 6CP4; 16-SEP-98.
DR      PDB; 1NOO; 08-MAR-96.
DR      PDB; 1PHA; 31-OCT-93.
DR      PDB; 1PHB; 31-OCT-93.
DR      PDB; 1PHD; 31-OCT-93.
DR      PDB; 1PHD; 31-OCT-93.
DR      PDB; 1PHE; 31-OCT-93.
DR      PDB; 1PHF; 31-OCT-93.
DR      PDB; 1PHG; 31-OCT-93.
DR      PDB; 1AKD; 19-NOV-97.
DR      PFMAM; PF00067; p450.1.
DR      PROSITE; PS00086; CYTOCHROME_P450; 1.
KW      Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW      3D-structure.
FT      INIT MET          0
FT      BINDING          357       357     HEME.
FT      CONFLICT         55        56     MISSING (IN REF. 3).
FT      CONFLICT        276       276     E -> Q (IN REF. 3).
FT      CONFLICT        361       361     H -> S (IN REF. 3).
FT      CONFLICT        407       407     D -> N (IN REF. 3).
FT      TURN            16        17
FT      HELIX           20        22
FT      STRAND          23        23
FT      TURN           28        29
FT      TURN           32        33
FT      HELIX           34        36
FT      HELIX           38        46
FT      TURN           48        49
FT      STRAND          53        56
FT      HELIX           58        60

```

FT	STRAND	62	65
FT	HELIX	68	76
FT	TURN	78	80
FT	STRAND	81	82
FT	TURN	83	84
FT	HELIX	90	95
FT	TURN	99	102
FT	TURN	105	107
FT	HELIX	108	119
FT	HELIX	121	145
FT	TURN	146	146
FT	STRAND	147	149
FT	HELIX	150	153
FT	TURN	154	156
FT	TURN	157	167
FT	TURN	168	168
FT	HELIX	171	185
FT	HELIX	193	213
FT	TURN	219	224
FT	TURN	225	225
FT	STRAND	227	228
FT	TURN	229	230
FT	STRAND	231	232
FT	HELIX	235	230
FT	TURN	251	252
FT	HELIX	253	266
FT	HELIX	268	276
FT	HELIX	278	291
FT	TURN	292	292
FT	STRAND	295	295
FT	STRAND	297	301
FT	STRAND	305	307
FT	TURN	308	309
FT	STRAND	310	312
FT	TURN	314	315
FT	STRAND	317	320
FT	HELIX	322	324
FT	TURN	325	327
FT	TURN	329	331
FT	TURN	335	336
FT	TURN	340	341
FT	TURN	349	350
FT	HELIX	353	355
FT	TURN	358	359
FT	HELIX	360	377
FT	STRAND	382	383
FT	TURN	385	386
FT	STRAND	391	392
FT	STRAND	396	396
FT	STRAND	398	399
FT	STRAND	403	405
FT	HELIX	408	410
SO	SEQUENCE	414 AA; 46538 MM; ECA610293AD6207 CRC64;	
Query Match 100.0%; Score 2180; DB 1; Length 414;			
Best Local Similarity 100.0%; Pred. No. 7, 4e-159;			
Matches	414; Conservativity	0; Mismatches	0; Indels
			Gaps 0;
OY	1	TTTETIOSNANLAPLPHPVPEHLVDFEDMYNPSNLSAGVOEAMAVLQESNVPDLVWTRCNG	60
Db	1	TTEETIOSNANLAPLPHPVPEHLVDFDMYNPSNLSAGVOEAMAVLQESNVPDLVWTRCNG	60
OY	61	GHWITRQOLIREAEDVRHFSSECPITPRAGEAYDITPTSMOPPEQORRALANQYVG	120
Db	61	GHWITRQOLIREAEDVRHFSSECPITPRAGEAYDITPTSMOPPEQORRALANQYVG	120
OY	121	MPVVDKLNRIQOELACSTLESILRPOGOCNFTEDYAEPPIRIFMLLAGLPREDIPLHLKYL	180
Db	121	MPVVDKLENRIQOELACSTLESILRPOGOCNFTEDYAEPPIRIFMLLAGLPREDIPLHLKYL	180
OY	181	TDQMTRPQSGMTFAAKKALVDYDLPIITEQRROKPGTDAISIVANGQVNGRPTISDEAKR	240



Query Match	Similarity	Score	DB 1: Length
Query Match	28.7%	380	393
Match Local Similarity	28.7%	Pred. No. 8.1e-22	
Matches 110: Conservative	71: Mismatches	182: Indels	20: Gaps

  

QY	31	PSNLASGVQEAAMAVLQESNVDPDVTWTRCGHMIATRGQLIREAYEDYRHFESCEPFI	90
QY <td>31</td> <td>PSNLASGVQEAAMAVLQESNVDPDVTWTRCGHMIATRGQLIREAYEDYRHFESCEPFI</td> <td>90</td>	31	PSNLASGVQEAAMAVLQESNVDPDVTWTRCGHMIATRGQLIREAYEDYRHFESCEPFI	90
DB <td>23</td> <td>PMADVGRV--GKKTILRDAG--PVVF--MNGWYLL--TRREDVLAALRNKVFESSKRALDP-</td> <td>74</td>	23	PMADVGRV--GKKTILRDAG--PVVF--MNGWYLL--TRREDVLAALRNKVFESSKRALDP-	74
QY <td>91</td> <td>EAGEAYDIFPIPSMDPEPQOFALANQVGMVVDKLENRIQELACSLIESLPQGCNF</td> <td>150</td>	91	EAGEAYDIFPIPSMDPEPQOFALANQVGMVVDKLENRIQELACSLIESLPQGCNF	150
DB <td>75</td> <td>-PGNPLPVPLALFDPDEPHTRRYRIILQPFSPALSKALPSLRHTVAMIDALAGGECGA</td> <td>133</td>	75	-PGNPLPVPLALFDPDEPHTRRYRIILQPFSPALSKALPSLRHTVAMIDALAGGECGA	133
QY <td>151</td> <td>TEDVPEPPIRIRFEMLLAGLPREDIDHKLVTLDQM----TRPGSMTFPAEAKALVDYLIP</td> <td>206</td>	151	TEDVPEPPIRIRFEMLLAGLPREDIDHKLVTLDQM----TRPGSMTFPAEAKALVDYLIP	206
DB <td>134</td> <td>MADLNLDPFPELVLTGYPLEDRDLTGKMAVYATMSDRPHATADVAARELLEYTLTA</td> <td>193</td>	134	MADLNLDPFPELVLTGYPLEDRDLTGKMAVYATMSDRPHATADVAARELLEYTLTA	193
QY <td>207</td> <td>ITEORRQKPGTFAISIVANGVNGRPITSDSEAKRMCGLLVGGLDVTYVNFLSFSMEFLAK</td> <td>266</td>	207	ITEORRQKPGTFAISIVANGVNGRPITSDSEAKRMCGLLVGGLDVTYVNFLSFSMEFLAK	266
DB <td>194</td> <td>MVAEERRRNPGRVLSQV---QIGEDPLSEIVLGLSHLLIAGLDVTVAANGFSLLEIAR</td> <td>250</td>	194	MVAEERRRNPGRVLSQV---QIGEDPLSEIVLGLSHLLIAGLDVTVAANGFSLLEIAR	250
QY <td>267</td> <td>SPENHOELIERPERIPAAACEELLR-RFSLVADGRILTSDFEHGVOLKKGDIILPOMLS</td> <td>325</td>	267	SPENHOELIERPERIPAAACEELLR-RFSLVADGRILTSDFEHGVOLKKGDIILPOMLS	325
DB <td>251</td> <td>RPQLAMLRDNPQKQIRVFTEIETIRLEPSPAVAPRVATPEVTGVGWTLPAGSPVRLCMAAV</td> <td>310</td>	251	RPQLAMLRDNPQKQIRVFTEIETIRLEPSPAVAPRVATPEVTGVGWTLPAGSPVRLCMAAV	310
QY <td>326</td> <td>GLDEENMCPMIVDPSRQKVSHTTGGHSHLCSGHLARRELIYVLKELTRIPDPSTAP</td> <td>385</td>	326	GLDEENMCPMIVDPSRQKVSHTTGGHSHLCSGHLARRELIYVLKELTRIPDPSTAP	385
DB <td>311</td> <td>NRDGSDAMSTDELVMDGKYNRHMGGGGPRHRCGLSHLARLETLTLVGWNLIPDPFEIAR</td> <td>370</td>	311	NRDGSDAMSTDELVMDGKYNRHMGGGGPRHRCGLSHLARLETLTLVGWNLIPDPFEIAR	370
QY <td>386</td> <td>--GAQIOHKSGIVSGVALPLWV 406</td> <td></td>	386	--GAQIOHKSGIVSGVALPLWV 406	
DB <td>371</td> <td>DYAPETIRFSPKSA-LKNLPLRW 392</td> <td></td>	371	DYAPETIRFSPKSA-LKNLPLRW 392	

  

RESULT	4
CPXM_BACME	STANDARD: PRT: 410 AA.
AC	Q06069.
DT	01-FEB-1995 (Rel. 31, Created)
DT	01-FEB-1995 (Rel. 31, Last sequence update)
DT	15-FEB-2000 (Rel. 39, Last annotation update)
DE	CYCHOXOME P450(MEG) (EC 1.14.99.-) (STEROID 15-BETA-HYDROXYLASE)
DE	(STEROID 15-BETA-MONOXYGENASE).
CN	CYP106A2.
OS	Bacillus megaterium.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OC	Bacillus/Staphylococcus group; Bacillus.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-ATCC 13368;
RA	MEDLINE: 94049677.
RA	Rauschenbach R., Isernhaugen M., Noeske-Jungblut C., Boldol W.,
RA	Slewert G.;
RT	"Cloning sequencing and expression of the gene for cytochrome
RT	P450meg, the steroid-15 beta-monoxygenase from Bacillus megaterium
RT	ATCC 13368.";
RL	Mol. Gen. Genet. 241:170-176(1993).
RL	[2]
RN	CHARACTERIZATION.
RP	STRAIN-ATCC 13368;
RC	MEDLINE: 79194051.
RA	Berg A., Ingelman-Sundberg M., Gustafsson M.;
RA	"Purification and characterization of cytochrome P-450meg.";
RL	J. Biol. Chem. 254:5264-5271(1979).
RL	[3]
RN	CHARACTERIZATION.
RP	STRAIN-ATCC 13368;
RC	MEDLINE: 82091079.
RA	Berg A., Ratfer J.J.;
RT	"Studies on the substrate specificity and inducibility of cytochrome
RT	P-450meg.";
RL	Biochem. J. 196:781-786(1981).
CC	-1- FUNCTION: HAS THE CAPACITY TO HYDROXYLATE CERTAIN STEROIDS IN THE
CC	15-BETA POSITION.

CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its use	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL: Z21972; CA979985.1; -	
DR	PIR: S32216; S32216.	
DR	PIR: S39924; S39924.	
DR	PFAM: PF00067; P450.1.	
DR	PRINTS: PR00359; B450.	
DR	PROSITE: PS00086; CYTOCHROME_P450.1.	
KW	Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.	
FT	BINDING 355 355 HEME (BY SIMILARITY)	
SO	SEQUENCE 410 AA; 46955 MW; 9EPD9CF2EE0F810B CRC64;	
Query Match	17.0%; Score 371; DB 1; Length 410;	
Best Local Similarity	30.1%; Pred. No. 4.1e-21;	
Matches	109; Conservative 53; Mismatches 180; Indels 20; Gaps	77
OY	63 WATRGQLREAYEDYRHPSE-----CPTRPAGAYDFIP-TSMDPPEQOFRALA 115	
Db	44 WNVFYEYDVKKRYLSDYKHKHSYKRRTTISVGTDESEGSPERKIQTSPDPHRRKRSLL 103	
OY	116 NOVGMPPVVDKLENTLOACLSLSLRQGQCNTEEDYAEFPPIRIFMLAGLPREDIP 175	
Db	104 AAATPPRSIQNNEPRIQETADELIGOMDGTIDIVASLASPLPIVAMDLMGVPSKDL 163	
OY	176 HLKYLTDQMPTRPDGSGMTFAE-----AKEALYDLPIPIEORRQKPGDAISIVANGOV 228	
Db	164 LFKKAVDTILFEPDEKQDEVDKLRQVAKETQYILPIYVQKRLNPADDIISDLKSEY 223	
OY	229 NGRPTTSDAKRMGILLVGLDVTYVNFLSFSM-EFLAKSPHROELTERPERIPACEE 287	
Db	224 DQEMTTDEEVYRTMLIGAGVETTSILLANSFYSLLYDDKENVYQELHENDLVPQAYBE 283	
OY	288 LLR-RFSLVADGRILTSDFEFGYQALKGQDILLQMLSGLDEREKACMHNDFSR-QKV 345	
Db	284 MLRFENFLIKLDRYKENDLLGVELKEDDSVVMMSAANMDEMEDEPFTNIHPNNK 343	
OY	346 SHTTGGHSGHLCGLHARRETIIVTLKEVLTPIIPDSIAPGAQIOHK---SGVSGVAL 402	
Db	344 KHLTNGCPHFCGLPRLARLLEAKIALTAFLKFKHLEAVPSQLEBNLTDSATGOTLSL 403	
OY	403 PL 404	
Db	404 PL 405	
RESULT 5		
YJIB_BACSU		
ID YJIB_BACSU	STANDARD: PRT; 396 AA.	
AC 034374:		
DT 15-FEB-2000 (Rel. 39, Created)		
DT 15-FEB-2000 (Rel. 39, Last sequence update)		
DT 15-FEB-2000 (Rel. 39, Last annotation update)		
DE PUTATIVE CYTOCHROME P450 YJIB (EC 1.14.-.-).		
GN YJIB.		
OS Bacillus subtilis.		
OC Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC Bacillus/Staphylococcus group; Bacillus.		
RN 11		
RP SEQUENCE FROM N.A.		
RA STRAIN:168;		
RA Rivoletta C., Soldo B., Lazarevic V., Joris B., Maue C., Karamata D.;		
RT "A 35.7 kb DNA fragment from Bacillus subtilis chromosome containing a		
RT putative 12.3 kb operon involved in hexuronate catabolism and a		
RT perfect catabolite-responsive element."		

Query Match	16.6%	Score 362	DB 1	Length 396
Best Local Similarity	30.1%	Pred. No. 1	9e-20	
Matches 99	Conservative 55	Mismatches 157	Indels 18	Gaps 7

RESULT	6
Y218_MYCTU	
ID	Y218_MYCTU
AC	053563;
DT	15-FEB-2000 (Rel. 39, Created)
DT	15-FEB-2000 (Rel. 39, Last sequence update)
DT	15-FEB-2000 (Rel. 39, Last annotation update)
DE	PUTATIVE CYTOCHROME P450 RV3518C (EC 1.14.-.-).
GN	RV3518C OR MTW023.25C.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN	[1]
RP	SEQUENCE FROM N.A.
RX	SRRA1-H37rv;
RX	MEDLINE; 98293987.
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.
RA	Gordon S.V., Eiglmeier K., Gass S., Barry C.E. III, Tekala F.,
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Hornsey T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

Query Match	16.5%	Score 360;	DB 1;	Length 398;
Best Local Similarity	28.2%	Pred. No. 2.7e-20;		
Matches	107;	Conservative	69;	Mismatches 165;
			Indels	34;
			Gaps	8;

RESULT	7		
CPXM_BACSU			
ID	CPXM_BACSU	STANDARD:	PRT: 405 AA.
AC	P27632;		
DT	01-AUG-1992 (Rel. 23, Created)		
DT	01-AUG-1992 (Rel. 23, Last sequence update)		
DT	15-FEB-2000 (Rel. 39, Last annotation update)		
DE	CYTCHROME P450 109 (EC 1.14.-.-) (ORF405).		
GN	CYP109.		
OS	Bacillus subtilis.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC	Bacillus/Staphylococcus group; Bacillus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
KC	STRAIN-W23;		



```
AC P14762;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450(BM-1) (EC 1.14.14.1).
GN CYP106.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-25.
RC STRAIN=ATCC 14581;
RX MEDLINE; 90089408.
RA He J.S., Ruettinger R.T., Liu H.-M., Fulco A.J.;
RT "Molecular cloning, coding nucleotides and the deduced amino acid
RT sequence of P-450BM-1 from Bacillus megaterium.";
RL Biochim. Biophys. Acta 1009:301-303(1989).
RN [2]
RP SEQUENCE OF 1-25 FROM N.A.
RX MEDLINE; 95355495.
RA He J.S., Liang Q., Fulco A.J.;
RT "The molecular cloning and characterization of BM1p1 and BM1p2
RT proteins, putative positive transcription factors involved in
RT barbiturate-mediated induction of the genes encoding cytochrome
RT P450BM-1 of Bacillus megaterium.";
RL J. Biol. Chem. 270:18615-18625(1995).
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X16610; CAA34612.1; -
CC EMBL: S79230; AAC60495.1; -
CC DR PIR; S07764; O4BS6M.
CC DR HSSP; P33006; 1CPT.
CC DR PFAM; PF00067; P450; 1.
CC DR PRINTS; PR00359; BP450.
CC DR PROSITE; PS00086; CYTOCHROME_P450; 1.
CC DR Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
CC KM BINDING 356 356 HEME.
CC FT SEQUENCE 410 AA; 47460 MW; C9AE293E76745387 CRC64;
SQ

Query Match 15.3%; Score 334.5; DB 1; Length 410;
Best Local Similarity 27.7%; Pred. No. 2.5e-18;
Matches 96; Conservative 57; Mismatches 170; Indels 23; Gaps 7;

63 WITRQGLIREAEDVYHFSSE--CPFIPRAGEAADFP---TSMDEPBOGQFALA 115
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
45 WNFQYEHVQVLSNYDFSSDGGRTTIFVGDNSKKKSTPTNLNLDPDRHKARSL 104
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
116 NOVYGMFVDKLENRIQELACSLIESLRPOGOCFTEDVAEPPIRIFMLAGLPEDIP 175
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
105 AAATPRLSKNWEPRIKQIADLVEALQKNSTINIVDLSPPSLVIALDFGVPAVDY 164
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
176 HLKYLDDQTRPGQSMFAAKE-----ALDYLIPIIEQRQKGTALSIANGQV 228
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
165 QFRKWDVILFQPDQERLEIEQEKORAGAEYQYLPIVIERKSNLSDDIISDLQAEV 224
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
229 NGRPITSDEAKRMGGLLVGLDPTVNFVLSFSMEFLAKPPEHROELIERPERIPACEE 287
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
225 DGEFTDEELVHATMLLGLGAVETTSALINMRYSLFYDKSKLSYSELRNRELAPKAVEE 284
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
288 LIR-RFSLVADGRLTSDYEFHGVQLKKQDQILLPQMLSGIDEREENACPMHVDFSR-QKV 345
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
DB 285 MLRYRPHISRRDRTVKODNELLGVKLRKGDVYIAMSACNMDETMPENPSVDIHRPTNK 344
OY 346 SHTFEGHSLCIGQHLARREIIVTLKEM/LR-----IPDFSIAP 385
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 345 KHLTFNGPHFCUGAPLARLEMKIILEAFLEAFSHIEPFEFDELEP 390
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
YME6_MYCTU
ID YME6_MYCTU STANDARD: PRT; 428 AA.
AC 050696;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-FEB-2000 (Rel. 39, Last annotation update)
DE PUTATIVE CYTOCHROME P450 RV2266 (EC 1.14.-.-).
GN RV2266 OR MYC1339.44C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE; 98295987.
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin W., Holtroyd S.,
RA Hornsby T., Jagers K., Krogan A., McLean J., Moule J., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z77163; CAB00969.1; -
CC DR HSSP; P33006; 1CPT.
CC DR TUBERCULIST; RV2266; -
CC DR PFAM; PF00067; P450; 1.
CC DR PRINTS; PR00359; BP450.
CC DR PROSITE; PS00086; CYTOCHROME_P450; FALSE_NEG.
CC DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
CC KM BINDING 379 379 HEME (BY SIMILARITY).
CC FT SEQUENCE 428 AA; 47824 MW; 76B1FC5AE348591 CRC64;
SQ

Query Match 15.3%; Score 334.5; DB 1; Length 428;
Best Local Similarity 26.0%; Pred. No. 2.0e-18;
Matches 116; Conservative 73; Mismatches 188; Indels 69; Gaps 17;

5 IQSNANLAPRPPIPHPEHLVDFDPMYNPSNLSAGVQEAVALQOE-----SNVDPDV 54
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
7 IATRVNQT-PPEVP---IADIEL-----GSLDEWALDDVDYRQDAFATLRREAPISF 54
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
55 WTR-----CNGGHWATRQGLIREAEDY---RH---FSS-----ECQFIPREA 92
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
55 WPTIELPGFVAGNGHWALTK-----YDDVFYASRHPDIFSSYPNITINDQPELAEPY 107
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
93 GEAYDFTPSMDPEQNGQFALANOVYGMFVDKLENRIQELACSLIESL---RPOGOCN 149
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
108 GSM-----IYLDPRIQRLRSLIVSRAFTPKVAVATEAAVDRHRLVSSMIANNPDRQAD 162
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
150 FTEDYAEPPPIRIFMLAGLPEDIPHLKYLTPDQMT---RPDSMPFAEKEALYD---Y 203
+ : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 163 LVSELAGPLQLQICDMMGIPKADHRIFFMTNVILGSGPDLATDDEFMOVSADIGAY 222
QY 204 LPIIEORRORPGTDAISIVANGVNGRPITSPDEAKRMCGILLVGLDVTYVNFLEFSMEF 263
Db 223 ATMLAEERRRVNHHDDLSSLVEAEVDESERLSSREIASFFLLVAGNETRNMTTHGYLA 262
QY 264 LAKSPHEQRELIERPERI-PACEBELLRPS-LVADGRILITSPDEFHGVLKKGDQILLP 321
Db 283 LSHRPEQRDRMWSDFGLAETAVEEIVRMASPVYVMRRITLQDIELRGTKMAAGDKVSLW 342
QY 322 QMISGLDERENRGMHVDSRQKVSHTTF-GHGSNLCLOGHILARREITVTLKEMLTFRPD 380
Db 343 YCSANRDESEFADPMTFDILARNRPNHIGFGGGAHFCGLGANLARREITRAVDELRRMPD 402
QY 381 FSIAPGAQIOHKSGIVSGVQALPLW 406
Db 403 V-VATEEPARLLSQFHIGIKTLPVTW 427

RESULT 11
CPYK_BACSU
ID CPYK_BACSU STANDARD; PRT; 410 AA.
AC 008469;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME P450 (EC 1.14.-.-).
GN CYP4 OR CYP107J1.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE: 97431495
RA Belitsky B.R., Gustafsson M.C.U., Somenshein A.L., von Wachenfeldt C.;
RT "An lrp-like gene of Bacillus subtilis involved in branched-chain
RT amino acid transport."
RL J. Bacteriol. 179:5448-5457(1997).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE: 97453479.
RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.;
RA Duesterhoeft A., Ehrlich S.D.;
RT "Sequence of the Bacillus subtilis genome region in the vicinity of
RT the lrp operon reveals two new extracytoplasmic function RNA
RT polymerase sigma factors Sigv and Sigz."
RL Microbiology 143:2939-2943(1997).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y11043; CAAT1937.1; -
CC EMBL: 093876; AAB80898.1; -
CC EMBL: 299117; CAB14615.1; -
CC DR HSSP: Q00441; IOXA.
CC DR SUBTLIST: BG11929; CYP4.
CC DR PFAM: PR00067; p450; 1.
CC DR PRINTS: PR00359; BP450.
CC DR PROSITE: PS00086; CYTOCHROME_P450; 1.
CC DR OXIDOREDUCTASE; Monooxygenase; Heme.
CC FT BINDING 359 HEME (BY SIMILARITY).
CC FT SEQUENCE 410 AA; 47384 MW; 035E98E58AA761AB CRC64;
```

Query Match

15.3%; Score 333; DB 1; Length 410;

```
Best Local Similarity 27.2%; Pred. No. 3, 2e-18;
Matches 86; Conservative 69; Mismatches 139; Indels 22; Gaps 5;

QY 104 DPREGQFRLANOVGMVPYVDLENRIQELACSLIESLRPOQOCFTEDYAPFPIRIF 163
Db 100 DPEDHNRKRLTVOKAFTTHRTITLQLEDKIOHTIADSLDKVOPKFMVLVDYAPPLPIYI 159
QY 164 MLTAGLPEEDIPMLKYLTDQMTRPDGSMTFAEAKAL-----VYLLPIIPORQ 213
Db 160 SEMIGIPLEDROKFRWWSQAI-----IDFSDAPEFLQENDHLGFEVETLESLVKKRR 213
QY 214 KPCTDAISIVANGVNGRPITSPDEAKRMCGILLVGLDVTYVNFLEFSMEFLAKSPHROE 273
Db 214 EPAGDILISALIQAESGTOI-STEEILYSMTLLIVAGHEFTVNLITMVTALMCHNDQLER 273
QY 274 LIRERPAPACEELLRFSLV--ADGRITSDYERHGVLKKGDQILLPQMLSGDERE 331
Db 274 LRQOPDLMSAIEALRFHSPVELTTRWTAEFFILHGQETKKKDYITISLASANDERK 333
QY 332 NACPMHVDSRQKVSHTTGHGSHLCLOGHILARREITVTLKEMLTFRIPPSIAPGAQIOH 391
Db 334 FPNADIFDIERKNNRHILATFGHGHFCLGAQLARLEKIAISTLLRRCNPQL- KGEKKQM 392
QY 392 K---SGIVSGVQALPL 404
Db 393 KWKGNFLMRALIELPI 408

RESULT 12
CPYK_SACER
ID CPYK_SACER STANDARD; PRT; 405 AA.
AC P33271;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 107B1 (EC 1.14.-.-) (P450C11B1).
GN CYP107B1.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 172-203 AND 302-324.
RC STRAIN=NRRL 2338;
RX MEDLINE: 92121109.
RA Andersen J.F., Hutchinson C.R.;
RT "Characterization of Saccharopolyspora erythraea cytochrome P-450
RT genes and enzymes, including 6-deoxyerythronolide B hydroxylase."
RL J. Bacteriol. 174:725-735(1992).
CC -1- FUNCTION: NOT KNOWN, PROBABLY INVOLVED IN THE CATABOLISM OF
CC OCTANE AND GUAIACOL. IT DISPLAYS A WEAK ACTIVITY IN THE O-
CC DEALKYLATION OF 7-ETHOXYCOUMARIN.
CC -----
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M83110; AAA26483.1; -
CC DR PIR: BA2606; B42606.
CC DR HSSP: Q00441; IOXA.
CC DR PFAM: PR00067; p450; 1.
CC DR PRINTS: PR00359; BP450.
CC DR PROSITE: PS00086; CYTOCHROME_P450; 1.
CC DR OXIDOREDUCTASE; Monooxygenase; Electron transport; Membrane; Heme.
CC FT BINDING 352 HEME (BY SIMILARITY).
CC FT SEQUENCE 405 AA; 45238 MW; 71C93CEC1FDC53FD CRC64;
```



```

Query Match          15.2% Score 331.5; DB 1; Length 405;
Best Local Similarity 25.6%, Pred. No. 4.1e-18;
Matches 107; Conservative 73; Mismatches 185; Indels 53; Gaps 10;

Oy 18 VPEHVF-D----EDMYNSNLSAGVQEAAMAVLQESNVPLVTWTRCNG--GHWIATRGOLIR 72
      ||| | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 6 VFDLIAFFDAFAQODKN-----RKARREPVORI--RTYNGLDAMIITRYEDVK 53
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 73 EAVEDYRHFSSCEPPIPREAGEAYDFIPTSM-----DPEBOROFK 112
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 54 QALLD-----PRIAKDGFRTQOIIEIKRLDAERBRPGSPFDLGPHMLNTDPDRTRLR 105
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 113 ALANOVGMPPVYDKLENRIQLAELSLIESLRQGOCNFTEYAERPPIRIEMLGLPEE 172
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 106 KLVNAFAPARVEGRPRJEOITDDLRLARGSEVDILDEFAPFLPTIVISELGVEDS 165
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 173 DIPHLKYTLDTMDTRPDGSGMTFAAKE--ALYUYLPIIEOROKGTPAISIVANGVWG 230
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 166 RDDDRSRMNTNLV--DGSOPEAOASVAMVYLLIELLAKKTETEGDDILLTLEAVEDG 223
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 231 RPITSDEAKRMCGLLVLGGDLTVTNFLFSMEFLAKSPENROELTERPERIPACEELLR 290
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 224 DRLSGELLAMFYLLLVGHGETTMNLIGNCVSLICNPDLQAALRNDSILPGALIEFLR 283
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 291 RSLVADG--RLTSDIERHGVLKKKGQIILLPOLSLGDEENACPMHVDSRKAVST 348
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 294 YESPANGFERTTAABAVRGDVVIREGELVWVALGAANNDRGRFPDPFRFTTRETGHV 343
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 349 TFGHSHTLCIGHLARRETIYTKMKLFRIPPSIA--PGAQIQKSGSVGSVALPL 404
      |||| | | | | : | : | : | : | : | : | : | : | : | : | : |
Db 344 AFGHIHCYGAALARLEQIVAGLIERFPLRLMAASFDDLRFMFVYLMRGLEKLVP 401
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 13
CPXE_STRGO STANDARD: PRT: 405 AA.
NC PI8326:
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450-SU1 (EC 1.14.-.-) (P450-CVA1) (CYP105A1).
GN CYP105A1 OR SUAC.
OS Streptomyces griseolus.
OC Bacteria; Filicutes; Actinobacteria; Actinobacteridae;
   Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-31.
RC STRAIN=ATCC 11796;
RX MEDLINE: 90264332.
RA Omer C.A., Lemstra R., Little P.J., Dean C., Tepperman J.M.,
   Leco K.J., Romesser J.A., O'Keefe D.P.;
RT "Genes for two herbicide-inducible cytochromes P-450 from
   Streptomyces griseolus.";
RL J. Bacteriol. 172:3335-3345(1990).
CC -1- FUNCTION: METABOLISM OF A NUMBER OF SULFONYUREA HERBICIDES.
CC -1- INDUCTION: BY HERBICIDE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL, M32238; AAA26823.1; -.
DR PIR, A35401; A35401.
DR HSSP, P23295; 2ROM.
DR PFAM, PF00067; P450_1.
DR PRINTS, PR00359; BP450.
DR PROSITE, PS00086; CYTOCHROME_P450_1.
KW Oxidoreductase; Monooxygenase; Electron transport; Heme.
```

```

FT      INT_MET      0      0      HEME (BY SIMILARITY).
FT BINDING      354 AA      354 AA      92AB36E064FD0B3E CRC64;
SQ      SEQUENCE      405 AA      44081 MW;
Query Match      15 1%; Score 330; DB 1; Length 405;
Best Local Similarity 26.3%; Pred. No. 5.4e-18;
Matches 104; Conservative 74; Mismatches 193; Indels 24; Gaps 11;

OY      31 PSNLSAGVQV--EAMAVLQESNVDPDLVMPFRCNGCH-WIATRGQLREAVEDYR----- 79
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      16 PSNRSCPQPLPFGGYAQLRPTPGPLRLRYLYDROKAMVYTKHEAKKLGGDPLSSNRDD 75
OY      80 HFSSECPFLP--REAGEAVYFIPTSMDDPEOROFALANOVGVMPVDKLENRIOELACS 137
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      76 NPAPSPREAVRESPQAF----IGLDPEHGTERRRMPTISEPTVRIKGMREVEEYVHG 131
OY      138 LIESLRPQG--QCNFEDVDAEPPIPIFMLAGLPREDIPHLKYLTDOMTRPGSMTPFAA 196
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      132 FLDENMLAGPTADLVSOFLAPVPSVYICRLGLGVPAADHEFFQDASKRLVQSTDAOSALTA 191
OY      197 KEALVDYLPIPIEQRKRGKGTDAI-SIVANGGVNCRPTSDAKRMCGLLVGLGDPVN 255
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      192 RNDLAGYLDGLTQGTQTEEGAGLVGALVADQLANE-IDRELISAMILLIAGHETAS 250
OY      256 FLFSMEFLAKSPDEHQELIERPERIPAAICELLRRFSL-VADGRILTSDYEFHGVDK 313
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      251 MTSLSVITLIDHPQOYALRLADRSILVPGAVELLRLAIADLAGRVATADIEVGHLIR 310
OY      314 KDDQLLLQMLSGLDEREACPMHVDPSQKYSHTTFGHGSLCGLHAREIITYLKE 373
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      311 AGEGVIVANSIANNRGTYVEDPDALDIHRSARHHLAFGFGVHQCLGQMLRLALEVILNA 370
OY      374 WLTRIPDFSIA-PGAQIQHKSQ-IVSGVOALPLVW 406
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      371 LMDRVPITRLANVPEDQLVLRPGTTIQGVNELPYTW 405

RESULT 14
YC56_MYCGU
ID YC56_MYCTU STANDARD: PRT: 405 AA.
AC Q11062: 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PUJATIVE CYTOCHROME P450 RV1256C (EC 1.14.-.-).
GN RV1256C OR MTCY50.26.
OS Mycobacterium tuberculosis.
OC Bacteria; Filumcutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV.
RX MEDLINE: 98295987.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RT Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```
CC or send an email to license@sb-stb.ch).
CC -----
CC EMBL: 277137; CAB00896.1; -.
CC DR HSSP: P33006; ICPT.
CC DR TUBERCULIST: RV1256C; -.
CC DR PRAM: PR00067; P450; 2.
CC DR PROSITE: PS00086; CYTOCHROME_P450; 1.
CC DR Hypothetical protein; Oxidoreductase; Monoxygenase; Heme.
CC FT BINDING 354 354 HEME (BY SIMILARITY).
CC SEQUENCE 405 AA; 44580 MW; 72DEA6CB688FA48 CRC64;

Query Match 14.7%; Score 320.5; DB 1; Length 405;
Best Local Similarity 29.5%; Pred. No. 2.8e-17;
Matches 102; Conservative 51; Mismatches 160; Indels 33; Gaps 9;

OY 62 HMTATGGLREAYEDYRHSSECPFIRREAGE-----ANDFTPTSM-DPEEQROFALAL 115
DB 46 YVYLSRADVWSAARDHQFFSS-AQGLTVNYGLEMIGLHDPVWMDPPVHTFEKLV 104
OY 116 NQVYGVPRVYDKLENRIQELACLISLIESLRPGQCNFTEDYAEPPRIRIFMLAGLPEDIP 175
DB 105 SRQFTTRQVETVEPYRKFEYVERLEKIRANGGDIYTELEKPLPSMVYVAHYLGVPEDWT 164
OY 176 HLRVLDQMTTP--DGSMTFA-EAKKALYDYLPIIEQRORPGTDAIS--IVANGVN 229
DB 165 QDFGWTQAIYVANAADGATGALDAVGSMAVYFTGLIERRTERPADDAISHLVAAVGDAD 224
OY 230 GRITIDSEAKRMGGLLVGLDVTNVNFLSMEFLAKSPERHOBELIERPERIPACCELL 289
DB 225 GDTAGTLLSLAFTFTWYTGNDVTGMLGSMPLLNHRPDKRLLDPPGIDPAVEELL 284
OY 290 RRRSLVAD-GRITIDSYEEFGVOLKKDQILPQMLSGLDERE-----NACPMHV 338
DB 285 RLSPVOGLARTTRTROYTIDDTIIPAGRRVLLYXGSANRBEROYGRDAELDYTRCPRNI 344
OY 339 DFRQKVSHTTEFGHSHLCGLQHLARREIIVTLKEMLTFRIPFSIA 384
DB 345 -----LTFSHGHNHCIGAAARQCRAVLTTELLARCPDEFA 381

RESULT 15
BIOT_BACSU STANDARD; PRT; 395 AA.
ID BIOT_BACSU
AC P53554;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, last sequence update)
DT 15-FEB-2000 (Rel. 39, last annotation update)
DE BIOTIN BIOSYNTHESIS; CYTOCHROME P450-LIKE ENZYME (EC 1.14.-.-).
GN BIOT OR CYPT07H.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98048467.
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequencing and functional annotation of the Bacillus subtilis genes
RT in the 200 kb rmb-dnaB region.";
RT Microbiology 143:3431-3441(1997).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-stb.ch/announce/
CC or send an email to license@sb-stb.ch).
CC -----
CC EMBL: U51868; AAB17462.1; -.
CC DR EMBL: AF008220; AAC00266.1; -.
CC DR EMBL: Z99119; CAB14997.1; -.
CC DR HSSP: Q00441; IOXA.
CC DR SUBTLIST: BG11528; BIOT.
CC DR PRAM: PR00067; P450; 1.
CC DR PRINTS: PR00359; BP450.
CC DR PROSITE: PS00086; CYTOCHROME_P450; 1.
CC DR Biotin biosynthesis; Oxidoreductase; Monoxygenase;
CC KW Electron transport; Heme.
CC FT BINDING 345 345 HEME (BY SIMILARITY).
CC SEQUENCE 395 AA; 44865 MW; E4AC3AF2637ACEIA CRC64;

Query Match 14.6%; Score 318.5; DB 1; Length 395;
Best Local Similarity 27.3%; Pred. No. 3.9e-17;
Matches 89; Conservative 65; Mismatches 133; Indels 39; Gaps 9;

OY 75 YEDYRHSSECPFIRREAGEAYDFIPTSMDDPRQRFALANQVGMVYDKLENRIQEL 134
DB 70 YODLSHVQNMMLFQNO-----PDHRLRLTLASGAFYPRPTTESYQPIIEF 115
OY 135 ACSLIESLRPGQCNFTEDYAEPPRIRIFML--LADLPREDIPHLKLYLDQMTTPGSGWT 192
DB 116 VNHLLDQVQKKKMEVISDFA--FPLASPVIANIIGVPEDREOLKEMASLIQ--TID 170
OY 193 FAKKEALYD-----YLPIIEQRORPGTDAISIVANGVNGRPITSDEAKRMG 242
DB 171 FTRSRKALTEGNIMAYQAAAYFEKELQKKRHHQODMISMLKGRKDK-LTDEEASLC 229
OY 243 GLLVGLDVTNVNFLSMEFLAKSPERHOBELIERPERIPACCELLRRFS-LVADGRIL 301
DB 230 ILAIAIGHETVVALISNSVLCILQHPQOLKLENDPLIGTAVEECIAYESPOTMARVA 289
OY 302 TSDYEEFGVOLKKDQILPQMLSGDERENMC---PMHVDPSRKVSHTTEFGHSHLC 358
DB 290 SEDIDICGVTIKQGEQVYL--LLGANRDPSTIFTNPVDVITRSPNHLRSHGHVCL 346
OY 359 GQHLARREIIVTLKEMLTFRIPFSIA 384
DB 347 GSSILARLEAQIAINTLLQRRPISLTLA 372
```

Search completed: October 4, 2000, 13:04:51  
Job time: 1682 sec



**THIS PAGE BLANK (USPTO)**



```

Db 157 D-----HRTFRWGSAGFLSTAETVAEENQAEQAYAVMGDLIDRRKREPDDLVLSAL 209
OY 224 ANGVNAGRPTTSDAKRMCGLLVGLDVTYVNFSLFSMEFLAKSPEROLIERPERIPA 283
Db 210 VQARDDODSDISBELDLALGLLVAGYESTTTOIADFYLLMTPELRROLLDPELIPS 269
OY 284 ACEELLRLRESL---VADGRLTSDYEFHGVOLKKGDQILLPOMLSGLDERENACPMHVD 340
Db 270 AVELTRWVPLDVGTAFFRPAVEDVTLRGVTLIRAGEPVLASTGCAANDQAOFPDADRIDV 329
OY 341 SRKVSHTTTFGHSNLCLOGLARRELYTLKEMLRIPDFSLA-PCAQIQHSG-IVSG 386
Db 330 DRPNQHLGFGHGHCLAPLARVELQVALLEVLQRLPGIRLGIPETQLRMSEGMILRG 389
OY 399 VQALPLVM 406
Db 390 PLELPVW 397

```

## RESULT 2

```

087605 PRELIMINARY: PRT: 416 AA.
ID 087605:
AC 087605:
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, last annotation update)
DE CYTOCHROME P450 MONOOXYGENASE.
GN PICK OR PICK.
OS Streptomyces violaceus (Streptomyces venezuelae), and
OS Streptomyces venezuelae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC15439;
RA BETLACH M.C., KEALEY J.T., BETLACH M.C., ASHLEY G.W., MCDANIEL R.;
RT "Characterization of the macroliide P450 hydroxylase from Streptomyces
RT venezuelae which converts narbomycin to picromycin.";
RL Biochemistry 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-S.VENEZUELA; STRAIN=ATCC15439;
RX MEDLINE: 99051447.
RA XUE Y., WILSON D., ZHAO L., LIU H.-W., SHERMAN D.H.;
RT "Hydroxylation of macroliactones YC-17 and narbomycin is mediated by
RT the pikC-encoded cytochrome P450 in Streptomyces venezuelae.";
RL Chem. Biol. 5:661-669(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-S.VENEZUELA; STRAIN=ATCC15439;
RX MEDLINE: 9844533.
RA XUE Y., ZHAO L., LIU H.-W., SHERMAN D.H.;
RT "A gene cluster for macroliide antibiotic biosynthesis in Streptomyces
RT venezuelae: architecture of metabolic diversity.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
DR EMBL: AF087022; AAC64105.1; -
DR EMBL: AF079139; AAC68886.1; -
DR HSSP: 000441: 10XA.
DR PFAM: PF00067; p450.1.
DR PRINTS: PR00359; BP450.
KW Monooxygenase.
SQ SEQUENCE 416 AA: 46038 MW: CAD6EBA0 CRC32:

```

Query Match 16.7%; Score 363.5; DB 2; Length 416;  
 Best Local Similarity 28.3%; Pred. No. 1e-21;  
 Matches 97; Conservative 61; Mismatches 164; Indels 21; Gaps 7;

```

OY 76 EDYRHSSCPPIPRAGAYDPIPTSMDEQORFALANOVGVVVDKLENRIQOLA 135
Db 72 KWR--NSTTPTLEADALNHNLES--DPRHTRRLKLVAREFTKRVKVELLRVOEIV 127

```

```

OY 136 CSLIESL--RPGQCNFTEDYAEPPPIRLFMILAGLPEEDIPHLKYLTDQMPRDSMTF 193
Db 128 DGLVAMLAAPDGRADLMESLAMPPLPTIVISELLGVEPBPRAAFVMTDAFVPRDDPAQ 187
OY 194 AEKKAELYDLIPIIDQRKRGPTDAIS-IVANGVNGRITSDAKRMCGLLVGLDT 252
Db 188 QTAAMSGYLSRLIDSKRQDDEDLALVRTSDEGSLTSEELGMHILLVAGHET 247
OY 253 VVNFSLFSMEFLAKSPEROLIERPERIPACCELLRLRSVLADRIITSDYEF----- 307
Db 248 TVNLINGMTALLSHPDQALALRADMTLLDGAVEEMLR-----YEGVESATIRFVEPY 302
OY 308 --HGVLKKGQDILLPOMLSGLDERENACPMHVDFSROKVSHTTFGHSNLCLOGLIARR 365
Db 303 DLDDGIVIPAGDVTLYVLAADHRTPERFPPHRRDTRDAGHLAFGHIHFCIGAPLARI 362
OY 366 EIVYTLKEMLRIPDS--IAPGAQIQHSGIVSQVALPLVM 406
Db 363 EARIAVALLERCPDLALDVPSELVWYPMPIRGKALPIRW 405

```

## RESULT 3

```

052544 PRELIMINARY: PRT: 396 AA.
ID 052544:
AC 052544:
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, last annotation update)
DE CYTOCHROME P450 MONOOXYGENASE.
OS Amycolatopsis mediterranei.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardiaceae; Pseudonocardiopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S699;
RA KIM C.G., YU T.W., FRYHLE C., HANDA S., FLOSS H.G.;
RL J. Biol. Chem. 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S699;
RA AUGUST P.R., TANG L., YOON Y.J., NING S., MUELLER R., YU T.W.,
RA TAYLOR M., HOFFMANN D., KIM C.G., ZHANG X., HUTCHINSON C.R.,
RA FLOSS H.G.;
RL Chem. Biol. 5:0-0(0002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S699;
RA AUGUST P.R., TANG L., YOON Y.J., NING S., MUELLER R., HUTCHINSON C.R.,
RA FLOSS H.G.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF040570; AAC01709.1; -
DR HSSP: 000441: 10XA.
DR PFAM: PF00067; p450.1.
DR PRINTS: PR00359; BP450.
KW Monooxygenase.
SQ SEQUENCE 396 AA: 44293 MW: BA875129 CRC32:

```

Query Match 16.5%; Score 360; DB 2; Length 396;  
 Best Local Similarity 28.8%; Pred. No. 1.8e-21;  
 Matches 116; Conservative 60; Mismatches 183; Indels 44; Gaps 14;

```

OY 27 DWYNPSNLGAGVQEMAVLQESNVPLVTRNGCH--WIAIRGQILRAYEDYRFSSE 84
Db 15 DKFDPRAVFDLSREH-RPLAKMYPD-----GHVGIVSSYELVREVLSDLR-FSHS 64
OY 85 CP---FIPREAGEAVDFIPT-----SMDPEQORFALANOVGVVVDKLENRIQ 132
Db 65 CEVGHFPPVTHQGV--IFTHPLIPGMFTHMDPEHTRRKLLTGFTYRRASRLIPRAE 121
OY 133 ELACSLIESLRPG-OCNFTEDYAEPPPIRIFMLAGLPEED---IPLKYLTDQMPR 187
Db 122 AVAAEOIEVWRAKGAADVMDFAKPLVLRMLGELVGLPEERDRVYPVAVTLIHADEAP 181

```

```

OY 188 DQSMTEFAAKKALDYLLPIITEOROKRGTDASIVANGOVNRPITSDAKRMCGLLV 247
Db 182 AEA---AAAYEACKFEFDEVIEREORPODILISSLVT-----EDLQTELRNVLTLF 233
OY 248 GGLPTVVFATFSFMEFLAKSPENHQELIEREPAPACEELLRRFSL--VADGRLHSDY 305
Db 234 AGYTEETBALATGVFALLHHTDOLALRAPEKIDAAIIEELRYLTNOHTYRTALEDV 239
OY 306 EFHGVOLTKGQDILLPOMLSGLDERENACPMHVDFSROKVSHTTFGHSHLCSOHLARR 365
Db 294 KLEGLLKKGGDTVVSLEPAARNRPAKGGCPARELDIEDTSGSHVAFGCIHQCLGONLARI 355
OY 366 ELIYTLKEMLTRIDFSTIA-PGAOIQHK-SCIVSGVALPPLW 406
Db 354 ELRAGFTALLRAPEELRLAVPADPEVRLRLKQSVSVYKLLPLSW 396

```

RESULT	4			
054302		PRELIMINARY;	PRT;	404 AA.
ID	054302			
AC	054302;			
DT	01-NOV-1996 (TEMBLrel. 01, Created)			
DT	01-NOV-1996 (TEMBLrel. 01, Last sequence update)			
DT	01-NOV-1999 (TEMBLrel. 12, Last annotation update)			
DE	CYTCHROME P450.			
GN	RAPN.			
OS	Streptomyces hygroscopicus.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
CC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NRRL 5491;			
RX	MEDLINE: 95372374.			
RA	SCHWECKE T., APARCIO J.F., MOLNAR I., KOENIG A., KHAM L.E.,			
RA	HAYDOCK S.F., OLITNYK M., CAFFEY P., CORTES J., LESTER J.B.,			
RA	BOEHM G.A., STAUNTON J., LEADLAY P.F.;			
RT	"The biosynthetic gene cluster for the polyketide immunosuppressant			
RT	rapamycin.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 92:7839-7843(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NRRL 5491;			
RA	MOLNAR I., APARCIO J.F., HAYDOCK S.F., EE KHAM L., SCHWECKE T.,			
RA	KOENIG A., STAUNTON J., LEADLAY P.F., STAUNTON J., LEADLAY P.F.;			
RL	Gene 0:0-0(0).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NRRL 5491;			
RA	APARCIO J.F., MOLNAR I., SCHWECKE T., KOENIG A., HAYDOCK S.F.,			
RA	EE KHAM L., STAUNTON J., LEADLAY F., LESTER J.B., BOEHM G.A.,			
RL	STAUNTON J., LEADLAY P.F.;			
RL	Gene 0:0-0(0).			
DR	EMBL: X86780; CAA60465.1; -			
DR	HSSP: 000441; 10XA.			
DR	PFAM: PF00067; P450; 1.			
DR	PRINTS: PR00359; BP450.			
SO	SEQUENCE 404 AA: 45071 MW: 05AB94DF CRC32:			

Query Match	15.9%	Score 346.5	DB 2	Length 404
Best Local Similarity	28.6%	Pred. NO. 2.3e-20		
Matches 102	Conservative 63	Mismatches 177	Indels 15	Gaps 9
QY	63	WATGSGLLREYEDYRHFSSCEPF---	IPRAGEAYDP-ITPSMDPPQQRFRALANOV	118
		:  : : : : :    :   :    : : : :		
Db	50	WLAASMEVAKAYFVDP--FSRSATLGKQDPRVLPALQDDPVIMLMDPPETHRLRAYATKA		108
QY	119	VCMPVVDKLENIIEACLSLIESLRPOG-QCFNTEDEYAPPRIFRIMLAGLPEEDIPHL		177
Db	109	LTSRMEALRPPTQEVADDLIDKMLAKGAPADLMEDFALPLFTIMICELLYGVIEQOTKF		168
QY	178	KYLTDQMTLRPDGSMT---FAEAKALYDVLPIIEQRORQPGTDAISIVANGQVNGRPIT		234

[illegible]

SEQ	FEATURE	START	END	ORIENTATION	QUAL
1	SEQUENCE	410 AA			
2	SEQUENCE	410 AA			
3	SEQUENCE	410 AA			
4	SEQUENCE	410 AA			
5	SEQUENCE	410 AA			
6	SEQUENCE	410 AA			
7	SEQUENCE	410 AA			
8	SEQUENCE	410 AA			
9	SEQUENCE	410 AA			
10	SEQUENCE	410 AA			
11	SEQUENCE	410 AA			
12	SEQUENCE	410 AA			
13	SEQUENCE	410 AA			
14	SEQUENCE	410 AA			
15	SEQUENCE	410 AA			
16	SEQUENCE	410 AA			
17	SEQUENCE	410 AA			
18	SEQUENCE	410 AA			
19	SEQUENCE	410 AA			
20	SEQUENCE	410 AA			
21	SEQUENCE	410 AA			
22	SEQUENCE	410 AA			
23	SEQUENCE	410 AA			
24	SEQUENCE	410 AA			
25	SEQUENCE	410 AA			
26	SEQUENCE	410 AA			
27	SEQUENCE	410 AA			
28	SEQUENCE	410 AA			
29	SEQUENCE	410 AA			
30	SEQUENCE	410 AA			
31	SEQUENCE	410 AA			
32	SEQUENCE	410 AA			
33	SEQUENCE	410 AA			
34	SEQUENCE	410 AA			
35	SEQUENCE	410 AA			
36	SEQUENCE	410 AA			
37	SEQUENCE	410 AA			
38	SEQUENCE	410 AA			
39	SEQUENCE	410 AA			
40	SEQUENCE	410 AA			
41	SEQUENCE	410 AA			
42	SEQUENCE	410 AA			
43	SEQUENCE	410 AA			
44	SEQUENCE	410 AA			
45	SEQUENCE	410 AA			
46	SEQUENCE	410 AA			
47	SEQUENCE	410 AA			
48	SEQUENCE	410 AA			
49	SEQUENCE	410 AA			
50	SEQUENCE	410 AA			
51	SEQUENCE	410 AA			
52	SEQUENCE	410 AA			
53	SEQUENCE	410 AA			
54	SEQUENCE	410 AA			
55	SEQUENCE	410 AA			
56	SEQUENCE	410 AA			
57	SEQUENCE	410 AA			
58	SEQUENCE	410 AA			
59	SEQUENCE	410 AA			
60	SEQUENCE	410 AA			
61	SEQUENCE	410 AA			
62	SEQUENCE	410 AA			
63	SEQUENCE	410 AA			
64	SEQUENCE	410 AA			
65	SEQUENCE	410 AA			
66	SEQUENCE	410 AA			
67	SEQUENCE	410 AA			
68	SEQUENCE	410 AA			
69	SEQUENCE	410 AA			
70	SEQUENCE	410 AA			
71	SEQUENCE	410 AA			
72	SEQUENCE	410 AA			
73	SEQUENCE	410 AA			
74	SEQUENCE	410 AA			
75	SEQUENCE	410 AA			
76	SEQUENCE	410 AA			
77	SEQUENCE	410 AA			
78	SEQUENCE	410 AA			
79	SEQUENCE	410 AA			
80	SEQUENCE	410 AA			
81	SEQUENCE	410 AA			
82	SEQUENCE	410 AA			
83	SEQUENCE	410 AA			
84	SEQUENCE	410 AA			
85	SEQUENCE	410 AA			
86	SEQUENCE	410 AA			
87	SEQUENCE	410 AA			
88	SEQUENCE	410 AA			
89	SEQUENCE	410 AA			
90	SEQUENCE	410 AA			
91	SEQUENCE	410 AA			
92	SEQUENCE	410 AA			
93	SEQUENCE	410 AA			
94	SEQUENCE	410 AA			
95	SEQUENCE	410 AA			
96	SEQUENCE	410 AA			
97	SEQUENCE	410 AA			
98	SEQUENCE	410 AA			
99	SEQUENCE	410 AA	</		

[illegible]

```
RESULT 6
OC 059819 PRELIMINARY; PRT; 407 AA.
AC 059819;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE CYTOCHROME P450 (EC 1.14.14.1).
GN OLEP.
OS Streptomyces antibioticus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95255619.
RA RODRIGUEZ A.M., OLANO C., MENDEZ C., HUTCHINSON C.R., SALAS J.A.;
RT "A cytochrome P450-like gene possibly involved in oleandomycin
RT biosynthesis by Streptomyces antibioticus.";
RL FEMS Microbiol. Lett. 127:117-120(1995).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; U37200; AAA92553.1; -.
DR HSSP; 000441; 10XA.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR PFAM; PF00067; p450; 1.
DR PRINTS; PR00359; BP450.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT BINDING 356 HEME (BY SIMILARITY).
SQ SEQUENCE 407 AA; 44957 MW; 04411C60 CRC32;
```

```
Query Match 15.8%; Score 344.5; DB 2; Length 407;
Best Local Similarity 28.1%; Pred. No. 3.4e-20;
Matches 104; Conservative 63; Mismatches 168; Indels 35; Gaps 11;
```

```
QY 60 GGHWITRQOLIREAVEDYRHFSSSEC---PTIPREAGEAVDFIPT-----SMDPEPQ 108
DB 50 GRALMTTRMSDARIYVIGDSR-FSTAATDPATPR-----MFTPEPDGVLADDPDH 101
QY 109 RQFRALANOVGMVVDKLENRIQELACSLIESLRPQO-CNFTEDYAEFPPIRIMLA 167
DB 102 TRRLRLVGKAFTRARVEEMRPVRSLYVSLDDMVAHGSPADVEFLAVFPVAVICELL 161
QY 168 GLPEEDIPHLKYLTDQM---TRPDGSMTFEAKKALYDYL---PIIEORRQKPGTDAIS 221
DB 162 GVPLEDRDLFRFESDAMLSTR---LTAETIORVOODEFVYMDGLVARDAPTEDLIG 217
QY 222 IVANGVNRPTISDAKRMCGILLVGLDITVYNFLSFSMEFLAKSPHROELIERPERI 281
DB 218 ALALATNDNDHLTKGEIVNMGVSLTAGHETSYNQTITNLVHLTLTKRKRESLVADPALV 277
QY 282 PACEELLRRFSVLADG---RILTSYEFHGVOLKKQDQILLPQMLSGLDERENACPMHY 338
DB 278 PAVEEMELKTYTPCVSGSVRVATEDELSTLYVRAGECVVIFNANDEVEFDIADDEL 337
QY 339 DESRQKVSHTTEGSHLCLGQHLARREITVTLKEMLTIRIPDSIA-PSAQIQHKSQ-IV 396
DB 338 DEFREHNPPIAHFGHAGHICAGIQGLRLEQELALSALVRRFPDLDAEPVAGLKWQGMIL 397
QY 397 SGVQALPLWV 406
DB 398 RGLEROIVSM 407
```

```
RESULT 7
OC 032460 PRELIMINARY; PRT; 411 AA.
AC 032460;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE ORF 10.
OS Actinomadura hibisica.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
```

```
OC Actinomycetales; Streptosporanginae; Thermomonosporaceae;
OC Actinomadura.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-P157-2.
RX MEDLINE; 97480928.
RA DAIRI T., HAMANO Y., IGARASHI Y., FURUMAI T., OKI T.;
RT "Cloning and nucleotide sequence of the putative polyketide synthase
RT genes for pradiacin biosynthesis from Actinomadura hibisica.";
RL Biosci. Biotechnol. Biochem. 61:1445-1453(1997).
DR EMBL; D87924; BAA23153.1; -.
DR HSSP; 000441; 10XA.
DR PFAM; PF00067; p450; 1.
DR PRINTS; PR00359; BP450.
SQ SEQUENCE 411 AA; 44860 MW; 98A628B6 CRC32;
```

```
Query Match 15.1%; Score 329; DB 2; Length 411;
Best Local Similarity 26.4%; Pred. No. 6.3e-19;
Matches 114; Conservative 61; Mismatches 199; Indels 58; Gaps 16;
```

```
QY 4 TIOSNNLALPLPHVPEHLVFPEDMYNP-----SNLSAGVQEMAVIQESNVDPDVTWTRCN 59
DB 9 TVDPRPDVTPAFPPRPD-----DPQPCERARLKASDPVAKVLP-----T 50
QY 60 GGH-WIATGQOLIREAVEDYRHFSSSEC---PTIPREAGEAVDFIPT-----MDPEP 107
DB 51 GDAWVYTVYADRVFTSD-RRESKEAVTRPGAPR-----LIPQROGSKSLVIMDPPE 102
QY 108 QROFRALANOVGMVVDKLENRIQELACSLIESLRPQO-QCNFTEDYAEFPPIRIMLL 166
DB 103 HPRMRKIVSRAETARVEEMRAHVRLDTSGFVDEHVEHGPADLIHLALPLPVIVICEM 162
QY 167 AGPREDIPHLKYLTDQMTRPDGSMTFEAKK-----ALTYDLIRIIOQRKPGTDAI 220
DB 163 LGVPRDPRFQDWTDRML-TTGAPALQADELTKAAVGRKGLAEIDAKTAAPRADLL 221
QY 221 SIYANGOVNRPTISDAKRMCGILLVGLDITVYNFLSFSMEFLAKSPHROELIERPER 280
DB 222 SLSTRHADQ-GLSEELLTFGKTLAAGYHTTAITHSVNHLREPSRYARLRDRDPG 280
QY 281 IPACEELLRRFSVLADG---RILTSYEFHGVOLKKQDQILLPQMLSGLDERENACPM 336
DB 281 IPAVAVELL-RYGOIGGAGAIRIAYEDVEVGTLVRAGEAVPIPLFNANRPEVFPADPE 339
QY 337 HDFSQKVSHTTEGSHLCLGQHLARREITVTLKEMLTIRIPDSIA-PSAQIQHKSQI 395
DB 340 ELDUGRTDNPHIALGHGIVYCLGAPLARLELOVLETIVERPALRLAIDADITWRPGL 399
QY 396 V-SGVQALPLWV 406
DB 400 AFARPDALPIAW 411
```

```
RESULT 8
OC 059723 PRELIMINARY; PRT; 406 AA.
AC 059723;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE CYTOCHROME P450 LIN (EC 1.14.14.1).
GN LINC.
OS Pseudomonas incognita.
OC Bacteria; Proteobacteria.
RN [1]
RP SEQUENCE FROM N.A.
RA ROP J.D., GUNSAUS I.C., SLIGAR S.G.;
RT Submitted (JAN-1994) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; U23310; AAA25810.1; -.
DR HSSP; P33006; 1CPT.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
```











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:35:13 ; Search time 76.18 seconds

(without alignments)  
128.722 Million cell updates/sec

Title: US-09-246-451-11

Perfect score: 2180

Sequence: 1 TFEETIOSNANIKAPLPHPVE.....IVSGVOALPLWDPATTKAV 414

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq.36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2133	97.8	405	1 Y04128	Pseudomonas putida
2	1185	54.4	446	1 Y04126	Bacterial and mamm
3	364	16.7	396	1 R60777	Mycimiclin IV hyd
4	332	15.2	408	1 R77867	S. clavuligerus OR
5	330.5	15.2	587	1 W33274	S. fradiae tylosin
6	330	15.1	406	1 R11349	Cytochrome enzyme
7	329	15.1	411	1 W54389	Actinomadura hibis
8	324.5	14.9	398	1 W11585	Streptomyces prist
9	322	14.8	403	1 R11350	Cytochrome enzyme
10	316	14.5	412	1 R38309	Sequence of the P4
11	313.5	14.4	410	1 R51368	Protein containing
12	296.5	13.6	404	1 R14724	6-hydroxylase enco
13	190	8.7	398	1 R47521	Vitamin D hydroxyl
14	160	7.3	422	1 W36128	Daunomycin C-14 hy
15	160	7.3	474	1 W36132	N-ferminal modifie
16	155	7.1	422	1 W00729	Dauorubicin 14-hy
17	144	6.6	518	1 W67616	A. nidulans phenyl
18	144	6.6	533	1 R15057	Cytochrome P450C25
19	142	6.5	494	1 R62825	Human steroid-21-h
20	130	6.0	512	1 W93216	Human cytochrome P
21	128	5.9	508	1 W35711	Chrysanthum flavon
22	128	5.9	512	1 R72365	Human auxillary cy
23	128	5.9	512	1 R33172	Human cytochrome P
24	128	5.9	516	1 W67617	P. chrysogenum phe
25	127.5	5.8	1169	1 R76544	Mitochondrial cyto
26	127	5.8	512	1 R72366	Human auxillary cy
27	127	5.8	512	1 W00652	Cytochrome P450A1
28	127	5.8	512	1 R93173	Human cytochrome P
29	125	5.7	524	1 P70577	Rat hepato-cytochr
30	123.5	5.6	523	1 R59291	Rat liver cytochro
31	122.5	5.6	898	1 P61030	Entire coded seque
32	122.5	5.6	898	1 P61082	Complete translati
33	122.5	5.6	899	1 P61056	Translation of pla

34	122.5	5.6	1144	1 P81334	Expression prod. o
35	122.5	5.6	1150	1 P81335	Expression prod. o
36	122.5	5.6	1150	1 P81337	Expression prod. o
37	122.5	5.6	1162	1 P81336	Expression prod. o
38	122	5.6	503	1 R72363	Human cytochrome P
39	122	5.6	503	1 R81464	Human derived cyto
40	122	5.6	503	1 R93170	Human cytochrome P
41	122	5.6	503	1 Y05202	Human CYP3A4 prote
42	122	5.6	512	1 R72364	Human auxillary cy
43	122	5.6	512	1 R93171	Human cytochrome P
44	121	5.6	514	1 W34539	Cytochrome P450 cc
45	120.5	5.5	493	1 R81467	Human derived cyto

## ALIGNMENTS

Post deleted

RESULT 1	Y04128	standard; Protein: 405 AA.
ID	Y04128:	
AC	11-JUN-1999	(first entry)
DE	Pseudomonas putida cytochrome P450 protein P450-cam.	
KW	Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;	
KW	oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;	
KW	bioremediation; environmental pollutant.	
OS	Pseudomonas putida.	
PN	W09908812-A1.	
PD	25-FEB-1998.	
PR	17-AUG-1998; U16979.	
PR	20-AUG-1997; US-056754.	
PA	(UYRP ) UNIV ROCHESTER.	
PI	Jones JP, Shimoji M;	
DR	WPI: 99-190131/16.	
DR	N-PSDB; X19926.	
PT	New P450 fusion proteins - comprising a portion of a bacterial	
PT	cytochrome P450 protein and a portion of a mammalian cytochrome P450	
PT	protein	
PS	Disclosure: Page 12-13; 51pp; English.	
CC	The present invention describes a fusion proteins comprising a portion	
CC	of a bacterial cytochrome P450 protein and also a portion of a mammalian	
CC	cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or	
CC	any compound having a carbon-hydrogen bond. The fusion protein can be	
CC	used for hydroxylating a compound to be oxidised. It can also be used in	
CC	the bioremediation of an environmental pollutant. Since the fusion	
CC	protein is soluble, it can be subject to structural elucidation by x-ray	
CC	crystallography for designing functional proteins. It can be readily	
CC	expressed in soil bacteria to facilitate bioremediation. The present	
CC	sequence represents Pseudomonas putida cytochrome P450 protein P450-cam	
CC	from the present invention.	
SO	Sequence 405 AA;	

Query Match 97.8%; Score 2133; DB 1; Length 405;  
Best Local Similarity 99.8%; Pred. No. 3.4e-209;  
Matches 404; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	10	NIAPLPVPELVDFDMYNSNL	SAGVQEMAVLQESNVVDLWTRCNGSHMATRGQ	69
DB	1	NIAPLPVPELVDFDMYNSNL	SAGVQEMAVLQESNVVDLWTRCNGSHMATRGQ	60
QY	70	LIREVEYDRHSSCPFPREAGEAYDFIP	SMPPPEORORALANOVGVVVDKLN	129
DB	61	LIREVEYDRHSSCPFPREAGEAYDFIP	SMPPPEORORALANOVGVVVDKLN	120
QY	130	RIQELAGSLIESLRQGCNFTEDYAEP	PIRIFMLAGLPREDIPHLKYLLDQMTFRDG	189
DB	121	RIQELAGSLIESLRQGCNFTEDYAEP	PIRIFMLAGLPREDIPHLKYLLDQMTFRDG	180
QY	190	SMTFAEAKALDYLLPIIEORQRKPG	TDALSIIVANGOVNGPRTISDAKRMCGLLVGG	249
DB	181	SMTFAEAKALDYLLPIIEORQRKPG	TDALSIIVANGOVNGPRTISDAKRMCGLLVGG	240

```

QY 250 LDTVNFSLFSMEFLAKSPENHROELIERPERIPACCELLRRFSLVADGRILTSYEEHG 309
DB 241 LDTVNFSLFSMEFLAKSPENHROELIERPERIPACCELLRRFSLVADGRILTSYEEHG 300
QY 310 VOLKRGDQDILLPOMLSGDERKNACPMHYDFSRQVSHHTFPGHSLCIGHLARREIV 369
DB 301 VOLKRGDQDILLPOMLSGDERKNACPMHYDFSRQVSHHTFPGHSLCIGHLARREIV 360
QY 370 TLKEWILTRIPDSIAPGAQIOHKSGIVSGVALPLVWDPATTKAV 414
DB 361 TLKEWILTRIPDSIAPGAQIOHKSGIVSGVALPLVWDPATTKAV 405

RESULT 2
ID Y04126 standard; Protein: 446 AA.
AC Y04126:
DE 11-JUN-1999 (first entry)
KW Bacterial and mammalian chimeric cytochrome P450 protein.
KW Bacterial, mammalian; cytochrome P450; chimeric; fusion protein;
KW oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;
KW bioremediation; environmental pollutant.
OS Synthetic.
PN W09908812-A1.
PD 25-FEB-1999.
PR 17-AUG-1998; U16979.
PR 20-AUG-1997; US-056754.
PA (UYRP ) UNIV ROCHESTER.
PI Jones JP, Shimoji M;
DR N-PSDB: X19916.
PT New P450 fusion proteins - comprising a portion of a bacterial
PT cytochrome P450 protein and a portion of a mammalian cytochrome P450
PT protein
PS Claim 24: Page 6-8; 51pp; English.
CC The present sequence is a fusion proteins comprising a portion of a
CC bacterial cytochrome P450 protein and also a portion of a mammalian
CC cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or
CC any compound having a carbon-hydrogen bond. The fusion protein can be
CC used for hydroxylating a compound to be oxidised. It can also be used in
CC the bioremediation of an environmental pollutant. Since the fusion
CC protein is soluble, it can be subject to structural elucidation by X-ray
CC crystallography for designing functional proteins. It can be readily
CC expressed in soil bacteria to facilitate bioremediation.
SQ Sequence 446 AA;

Query Match 54.4%; Score 1185; DB 1; Length 446;
Best Local Similarity 57.6%; Pred. No. 1.8e-112;
Matches 253; Conservative 33; Mismatches 99; Indels 54; Gaps 9;

QY 10 NLAPRPVPHVPEHLVDFPMYNPSNLACVORAMAVLOESNVVDLWVTCNGSHWATRGQ 69
DB 1 NLAPRPVPHVPEHLVDFPMYNPSNLACVORAMAVLOESNVVDLWVTCNGSHWATRGQ 60
QY 70 LIREAYEDYRHFSSECPPIPREAGEAYDFIPTSMDPBQORFALANOVGMPPVVDKLEN 129
DB 61 LIREAYEDYRHFSSECPPIPREAGEAYDFIPTSMDPBQORFALANOVGMPPVVDKLEN 120
QY 130 RIQELACSLIESLRPQGCNTEDYAERPIRIFMLLAGLREEDIPHLKYLTDQTRPDG 189
DB 121 RIQELACSLIESLRPQGCNTEDYAERPIRIFMLLAGLREEDIPHLKYLTDQTRPDG 180
QY 190 SMTPAEAKALDYLIPIIEBOROKPCT-----DAISIVANGVGNRP--ITSSEAKRM 241
DB 181 SMTPAEAKALDYLIPIIEBOROKPCT-----DAISIVANGVGNRP--ITSSEAKRM 240
QY 242 GCLLVGLDITVNFSLFSMEFLAKSPENHROELIERPERI-----PA 283
DB 241 AVDLFGAGTETTTSTLRVALLLKLNHPVTAKVGEIERVIGRNKSPCMQDSHMPYIDA 300
QY 284 ACEELLRRFSLVADG--RIILTSYEEFGVQLKGDQILL-----POMLSG- 326
DB 284 ACEELLRRFSLVADG--RIILTSYEEFGVQLKGDQILL-----POMLSG- 326

```

```

DB 301 VHEVQRIYIDLPIPSLPAVNTCDIKERNYLIPKGTITLISLTSVLHDKREPPNEMEDPH 360
QY 327 --LDERKNACPMHYDFSRQVSHHTFPGHSLCIGHLARREIYTLKEWILTRIPDSIA 384
DB 361 HFLDDEGNN-----FKRSKY-FMFSAGKRICVGEALAGNELFLITSLIQNFMKSLV 412
QY 385 PEAQIOHKSGIVSGVALP 403
DB 413 DPKNID-TTPPVNCGFASVP 430

RESULT 3
ID R60777 standard; Protein: 396 AA.
AC R60777:
DE 21-JUN-1995 (first entry)
KW Mycinamicin IV hydroxylating protein.
KW Mycinamicin; hydroxylase; macrolide; antibiotic; Micromonospora;
KW Micromonospora griseorubida.
OS Micromonospora griseorubida AT11725CN3.
PN J06253853-A.
PD 13-SEP-1994.
PR 09-MAR-1993; 047638.
PR 09-MAR-1993; JP-047638.
PA (ASAH ) ASAH KASEI KOGYO KK.
DR WPI: 94-328997/41.
DR N-PSDB: Q73674.
PT DNA encoding a protein having mycinamicin IV hydroxylating
PT activity - for prodn. of mycinamicin, a macrolide antibiotic
PS Claim 1: Page 12-14; 23pp; Japanese.
CC The amino acid sequence of a protein having mycinamicin IV hydroxylase
CC (MH) activity. The gene encodes a protein of 396 a.a. The DNA was
CC obtained from the macrolide antibiotic-producing bacterium Micromonospora
CC griseorubida A11725CN3/pTYS507. The gene was isolated from the plasmid
CC pTYS507. The protein encoded by this plasmid can be used to produce
CC mycinamicin IV in pTYS507-deficient Micromonospora strains.
SQ Sequence 396 AA;

Query Match 16.7%; Score 364; DB 1; Length 396;
Best Local Similarity 28.5%; Pred. No. 7.9e-29;
Matches 105; Conservative 57; Mismatches 168; Indels 38; Gaps 8;

QY 63 WIATRGQILREAYEDYRHFSSECPPI-----PREAGEAYDFIPTSMDPBQORFRA 113
DB 43 WLWTR-----YEDYRAVLGDGRFVRGPMTRDEPFRPRPMVVGGLSMPPHESRLRR 95
QY 114 LANOVGMPPVVDKLENRIQELACSLIESLRPQGCNTEDYAERPIRIFMLLAGLREE 172
DB 96 LVVNAFTARRRESLRPRAREIAHLELVDMATGQPADLVAMFARQLPYRVICELLGVPESA 155
QY 173 DIPHLKYLTLDQTRPDGSM-----TFPAEKALDYLIPIIEBOROKPGTDAISIV 223
DB 156 D-----HDFTRKSGAFSLTAEYTAEMQDAEQAYAGDILDRKRKPTDVLVSAL 208
QY 224 ANGVNRPITSDAKRKCGLLVGLDITVNFSLFSMEFLAKSPENHROELIERPERIPA 283
DB 209 VQARGQDQSLSEQELDLALGIVAGVSTTQIADFYLLMTMPRELKROLDLRPELIS 268
QY 284 ACEELLRRFSLVADGRILTSYEEFGVQLKGDQILLPOMLSGDERKNACPMHYDF 340
DB 269 AVEELTRVPLGVGTAFPRYAVEVTLRGVYIRAGEPLASTGANRQDAQFPDADRIDV 328
QY 341 SRQVSHHTFPGHSLCIGHLARREIYTLKEWILTRIPDSIA--PEAQIOHKSG-IVSG 398
DB 329 DRTPNQHLGFHGVNHCGLAPLAVEILOVALLEVLLQRLPGIRIGIPETOLWSEGMLRG 388
QY 399 VQALPLVW 406
DB 389 PLELPVW 396

RESULT 4

```

R77867  
ID R77867 standard; Protein: 408 AA.  
AC R77867;  
DE 13-NOV-1995 (first entry)  
DE S. clavuligerus ORF10 product.  
KM Clavulanic acid; clavulinate; antibiotic; beta-lactamase-inhibitor.  
OS Streptomyces clavuligerus.  
PN CA2108113-A.  
PD 09-APR-1995.  
PF 08-OCT-1993; 108113.  
PR 08-OCT-1993; CA-108113.  
PA (UYAL-) UNITV ALBERTA.  
PI Aldoo KA, Jensen SE, Paradkar AS;  
DR N-PSDB: Q91580.  
DR Clavulanic acid biosynthesis enzymes and corresp. DNA - useful for  
PT biosynthesis of the antibiotic in Streptomyces hosts which do not  
PT naturally produce clavulanate  
PS Claim 32; Fig.19; 41pp; English.  
CC A 15 kb fragment S. clavuligerus NRRL 5741 genomic DNA (Q91580),  
CC extending downstream from pcbc, included 10 ORFs encoding the  
CC enzymes required for clavulinate biosynthesis. The ORF10  
CC product (R77867) showed high similarity to cytochrome P450-type  
CC enzymes from other Streptomyces spp.  
SQ Sequence 408 AA;

Query Match 15.2%; Score 332; DB 1; Length 408;  
Best local similarity 29.0%; Pred. No. 1.5e-25;  
Matches 122; Conservative 59; Mismatches 185; Indels 54; Gaps 18;

QY 16 PVPHPHLYVDFDMYNPNSAGVOEAMVLOESNVPDLVWTCNGSH-WIATRGQLIREA 74  
DB 14 PAVPHRHVCPVD---PPQIAGLRBQKAKASRYT-----LM---DSQVWLVTHAGARAV 62  
QY 75 YEDYRHFS-SECPFR-----REAGEAYDFIPTSMDPPEORQFRA-----LANQV 118  
DB 63 LGDRFRTAVTSAPGRPMILRTSQLVRANPESASF--RMDPDQHSRLSMLTRDFLARRA 120  
QY 119 VGM-PVNVKLEKRIQELACLISLESIRPQGCNFTEDYAEPPFIRIRFMLAGLPEDDIPHL 177  
DB 121 EALRAVNEHL--LDEILGGLVKGEP--VDLVAGLTIPVPSRVITLFLFGAGDDRREFI 174  
QY 178 K-----YLTDMQTRPDGSMTEFAEKALYDYLPIIEORROKGTALISIVANGVNGRP 233  
DB 175 EDRSAVILDRGTYRE---QVAKARDELQCYLRELYBERKENGTDLISLVLDQVRPHLU 231  
QY 234 TSDAARRMCGILLVGLDVTNVNLFSSMEFLAKSPENROELIERPERIPACEELLRRFS 293  
DB 232 RVEEAVPMCRLLLVAGHGTTTQASLSTLSTLTDPELAGRLTEDPALLPKAVEELLRFHS 291  
QY 294 LVADG--RLTSDYEFHGVOLKKGDQILLPOMLSGLDERKKACPMHNVDSRQKVSHTTFG 351  
DB 292 IVQNGIARAAYEDVQLDVYLIRAGGVVLSAGNRDEIVLPDPVRVDYDRARRHLLAFG 351  
QY 352 HGSNCLGSOHLAR---REIIVTLKEMLTGRIPDESIA-PCAOIOHKSQIYS-GVQALLPLW 406  
DB 352 HGMHOCLOGMARVLEELTAAVLKRM---PGARLAVPPEELDFRHEVSSIGLALPVTW 408

RESULT 5  
ID W33274  
AC W33274 standard; Protein: 587 AA.  
DE 17-FEB-1998 (first entry)  
DE S. fradiae tylosin biosynthesis gene product tyIH.  
KM Tylosin; biosynthetic gene product; production; antibiotic; tyIH.  
OS Streptomyces fradiae.  
PN US5672497-A.  
PD 30-SEP-1997.  
PF 21-DEC-1995; 575843.  
PR 12-MAY-1989; US-351350.  
PR 21-MAR-1986; US-842330.

PR 25-JUL-1986; US-890670.  
PR 24-FEB-1987; US-018237.  
PR 06-AUG-1991; US-742222.  
PR 28-JUL-1993; US-107232.  
PR 17-FEB-1994; US-198672.  
PR 21-DEC-1995; US-575843.  
PA (ELIL) LILLY & CO ELI.  
PI Cox KL, Fishman SE, Hersberger CL, Seno ET;  
DR N-PSDB: T58686.  
DR DNA encoding Streptomyces fradiae tylosin biosynthesis gene products  
PT - for increasing tylosin production in Streptomyces spp.  
PS Claim 20; Columns 21-38; 38pp; English.  
CC The present sequence is the Streptomyces fradiae tylosin  
CC biosynthetic gene product tyIH, useful to increase the production  
CC of the antibiotic tylosin in Streptomyces spp.  
SQ Sequence 587 AA;

Query Match 15.2%; Score 330.5; DB 1; Length 587;  
Best local similarity 29.3%; Pred. No. 3.8e-25;  
Matches 105; Conservative 62; Mismatches 170; Indels 21; Gaps 11;

QY 63 WIATRGQLIREAYEDYR--HFSSECPFIREAGEA-YDFIPTSMDPPEORQFRLANQV 118  
DB 143 WLISRODHVRLALDAPRVSIH-PAKLPRLSPSDGEASRSLLTIDPPDHGALRGHFIPE 201  
QY 119 VGMPPVVDKLEKRIQELACLISLESIRPQGCNFTEDYAEPPFIRIRFMLAGLPEDDIPHL 177  
DB 202 FELIRVRELRPEVSEIVGLDLDLDTARGLDEADLDLAFALPMATQVYICRLDIPYEDRDFV 261  
QY 178 KYLTQMTPEPDGSMTEFAEKALYDYLPIIEORROKGTCTDAI--SIYANGVNGRPITS 235  
DB 262 QERTQADTPRAAGEALELELRDLRLISGKGTRESGDMLGSVA--OARGGGLSH 319  
QY 236 DEAKRMCGILLVGLDVTNVNLFSSMEFLAKSPENROELIERPERIPACEELLRRFSIV 295  
DB 320 ADVLNAVYLLAAGHETTASMTVMGVLYLQHPTRAMRELTVPAGLLPAGVDBLLAYLS-I 378  
QY 296 ADG--RLTSDYEFHGVOLKKGDQILLPOMLSGLDERKKACPMHNVDSRQKVSHTTFGCG 353  
DB 379 ADGLRRSATADLEIDGHTIRAGDGLVFLLAANRDEAVSEPEADRIHSARRHAAFGYG 438  
QY 354 SHLCIGOMLARREIIVTLKEMLTGRIPDESIA-----GAOIOHKSQIYS-GVQALLPLW 406  
DB 439 PHOCIGQMLARMELEVALGAVLERLP--ALRPPTVAGRLKSDSA-VFGVTELPVAM 493

RESULT 6  
ID R11349  
AC R11349 standard; Protein: 406 AA.  
DE 05-JUN-1991 (first entry)  
DE Cytochrome enzyme P450S01.  
KM Cytochrome P450; P450S01; P450S02; herbicide resistance.  
OS Streptomyces griseolus.  
PN W09103561-A.  
PD 21-MAR-1991.  
PF 27-AUG-1990; 004785.  
PR 11-SEP-1989; US-405605.  
PR 12-JAN-1990; US-464499.  
PR 23-AUG-1990; US-569781.  
PA (DUBO) DU PONT DE NEMOURS CO.  
PI Dean C, Harder PA, Leto KJ, Lichtner FT, Odell JT;  
PI O'Keefe DP, Omer CA, Komesser JA;  
DR WPI: 91-102077/14.  
DR N-PSDB: Q11126.  
PT DNA encoding cytochrome P450 enzymes - and electron donating  
PT iron-sulphur proteins, used to confer herbicide resistance to  
PT plants and microorganisms  
PS Claim 13; page 151; 224pp; English.  
CC This cytochrome P450 enzyme, P450S01 is expressed alongside the iron  
CC sulphur protein Fes-B, by a DNA sequence contained in a recombinant

CC plasmid. Host Streptomyces species are transformed with the plasmid  
CC and are used to coat a plant seed to transform the plant. The res-  
CC ultant transformants are resistant to herbicides.

CC See also Q11127.  
SQ Sequence 406 AA;

Query Match 15.1%; Score 330; DB 1; Length 406;  
Best Local Similarity 26.3%; Pred. No. 2,4e-25;  
Matches 104; Conservative 74; Mismatches 193; Indels 24; Gaps 11;

QY 31 PSNISAGVQ--EAMAVIOESNVPDLVWTRNGGH-WIATRGOLIREAYDYR----- 79  
DB 17 PSNRSCPYQLPDGYAQLDRTGCPILHRTVLYDGRQAMVYTKHKAARKLLGDPLSSNRTDD 76  
QY 80 HFSSECPRI--REAGEAYVDPIPTSMDDPEOROFALANOVGVMDVKLEKRIQELACS 137  
DB 77 NEPAISPRFEAVRESPQAF---IGLDPRHEGTRRRMTISEFTYRRIGMRPEVEVYHG 132  
QY 138 LIESLRPGG-OCNFTEDYAEPPPIRIFMLLAGLPREDIPHLKYTLDDMTRPDGSMTFAEA 196  
DB 133 FLDEMLAGPPADLVSOALPVPMSVTCRLGVPYADHEFPDASKRLVQSTDAQSALTA 192  
QY 197 KEALDYILPIITEQRQRKPGTDAI-SIVANGQVNGRPITTSBEAKRMCGLLLVGGLDTYVN 255  
DB 193 RNDLAGYLDGLITQFOTEPGAGLVGALVADOLANGE-IDREELLSTAMLLIAGHETAS 251  
QY 256 FLFSMEFLAKSPENHROELIERPERIPAACELLRFSL--VADRIITLSQYEFHGYOLK 313  
DB 252 MTSLSVITLLDHPHOYAAALRADRSVLPGAVBELLYLAIDAGGAVTADIEVEGHILIR 311  
QY 314 KGDQILLPQMLSGDERKNACPMHVFESRQKVSHTTFGSHLCLGQHLARREIIVTLKE 373  
DB 312 AGEYIIVNSIANDRGVYEDPDALDIHRSARHHLAFGEVHGQCGQLARLELEVIINA 371  
QY 374 WLTRIPDSIA-PCAQIOHKSG-IYSGVALPLVW 406  
DB 372 LMDRVPTRLAVPVEQVLVRGTTIGVNELPVTW 406

RESULT 7  
W54389  
ID W54389 standard; Protein; 411 AA.  
AC W54389;  
DE 18-AUG-1998 (first entry)  
DR Actinomadura hibisca polyketide synthase protein 10.  
KW Multienzyme; infection; fungi; yeast; gram positive bacteria; virus;  
KM dihydrobenzo(a)naphthacenequinone aglycon; antibiotic; Pradimicin.  
OS Actinomadura hibisca.  
PN MO9811230-A1.  
PD 19-MAR-1998.  
PE 13-SEP-1996; WI4791.  
PR 13-SEP-1996; WO-014791.  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
PI Dairi T, Oki T;  
DR WPI: 98-207391/18.  
DR N-PSDB: V26609.  
DR P-PSDB: W54389, W54390.  
PT Actinomadura polyketide synthase genes - useful for preparation of  
PT Pradimicin  
PS Disclosure: Page 55-56; 71pp; English.  
CC The Actinomadura hibisca polyketide synthase proteins W54380-W54390 form  
CC a multienzyme complex. The enzyme is used for the biosynthesis of a  
CC dihydrobenzo(a)naphthacenequinone aglycon preferably a Pradimicin which  
CC is an antibiotic useful against systemic fungal infections caused by  
CC Candida albicans, Aspergillus fumigatus and Cryptococcus neoformans. It  
CC is also active against a wide variety of fungi and yeasts, some  
CC Gram-positive bacteria and viruses.  
SQ Sequence 411 AA;

Query Match 15.1%; Score 329; DB 1; Length 411;  
Best Local Similarity 26.4%; Pred. No. 3.1e-25;

Matches 114; Conservative 61; Mismatches 199; Indels 58; Gaps 16;  
QY 4 TIQSANLALPRPHVPEHLVDFQKYNP-----SNISAGVQAMAVIOESNVPDLVWTCN 59  
DB 9 TVDPRDVTAPAFPERPD-----DFQPPCEHARIRASDPVAKVLP-----T 50  
QY 60 GGH-WIATRGOLIREAYDYRHFSSFC--PIPREAGEAYDPIPTSMDDPEOROFALANOVGVMDVKLEKRIQELACS 107  
DB 51 GDHAMVYIRVADRVYTSQD-RFESKEAVTRGCAFR-----LIPMOGSKSLVMDPE 102  
QY 108 QORFALANOVGVMDVKLEKRIQELACSLSIESLRPGG-OCNFTEDYAEPPPIRIFMLL 166  
DB 103 HTMRKIVSRAFTARRVGMRAHVADLTSGFVDEKVEHGPADLIALDLPLPVVICEM 162  
QY 167 AGLPREDIPHLKYTLDDMTRPDGSMTFAEAKE-----ALDYILPIITEQRQRKPGTDAI 220  
DB 163 LGVPEDEPRFDQWTDNRML-TIGAPALAODEIKAAVGRGLGYLAELLIDARTAAPADLL 221  
QY 221 SIVANGQVNGRPITTSDEAKRMCGLLLVGGLDTYVNFLSFSMEFLAKSPENHROELIERPER 280  
DB 222 SLISRAHADG-GLSFEELLTFCMTLLAGYHTTTAITHSYHLLRERSRRLREDPSG 280  
QY 281 IPACEELLRRFSLVADG----RITSDYEFHGYOLKKGQDILLPQMLSGDERKNACPM 336  
DB 281 IPAAVEELL-RYGOIGGAGAIRIAVEDVEVGLVFRAGEAVIPLFNANRBDPEVFADPE 339  
QY 337 HVDSPROKVSHTTFGSHLCLGQHLARREIIVTLKEWLTPIRPSIA-PCAQIOHKSGI 395  
DB 340 ELIDGRITNPRIALHGHGICLIGAPLARLEQVVLVETLVERTPALRIALIDDAITWRGCL 399  
QY 396 V-SGVQALPLVW 406  
DB 400 AFARDALPIAM 411

RESULT 8  
W11585  
ID W11585 standard; Protein; 398 AA.  
AC W11585;  
DE 02-APR-1997 (first entry)  
DR Streptomyces pristinaespiralis snbF gene product.  
KW Streptogramin B; antibiotic; biosynthesis; pristinamycin;  
KM virginiamycin; pipercolic acid; cyclodeamination; papa; snba; snbF;  
KW p1pA; 3-hydroxy-picolinic acid; hydroxylation.  
OS Streptomyces pristinaespiralis.  
PN MO9601901-A1.  
PD 25-JAN-1996.  
PE 04-JUL-1995; F00889.  
PR 08-JUL-1994; FR-008478.  
PA (RHON ) RHONE-POULENC RORER SA.  
PI Barriere JC, Blanc V, Blanche F, Crouzet J;  
PI Debussche L, Paris JM, Thibaut D, Bamas-Jacques N;  
PI Dutruc-Kossel G, Famechon A;  
DR WPI: 96-097631/10.  
DR N-PSDB: J58555.  
PT New streptogramin B derivs. useful as antibiotics - produced by new  
PT mutants of Streptomyces having altered genes for streptogramin B  
PT biosynthesis  
PS Example 1; Page 113-114; 146pp; French.  
CC The papa gene of S.pristinaespiralis is involved in the biosynthesis  
CC of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for  
CC pristinamycin 1A. Upstream of the papa gene, on the complementary  
CC strand, is the snba gene coding for 3-hydroxy-picolinic acid-AMP lyase.  
CC The region between these two genes was sequenced and two open reading  
CC frames were identified. The first (p1pA) decodes to an amino acid  
CC sequence with homology to ornithine cyclodeaminase from Agrobacterium  
CC tumefaciens. The p1pA gene product is likely to catalyse the cyclo-  
CC deamination of lysine, leading to production of pipercolic acid.  
CC Mutations in the p1pA gene were shown to affect pipercolic acid  
CC synthesis but not the synthesis of 3-hydroxy-picolinic acid. The second  
CC open reading frame (snbF) could be decoded to give a product with  
CC homology to hydroxylases of the cytochrome P450 type.  
CC Disruption of the p1pA and snbF genes can be used to produce



CC strains of *S. pristinaespiralis* which are unable to produce the  
 CC antibiotic pristinamycin I but which may be able to produce new,  
 CC modified forms of it.  
 SQ Sequence 398 AA;

Query Match 14.9%; Score 324.5; DB 1; Length 398;  
 Best Local Similarity 29.0%; Pred. No. 8.5e-25;  
 Matches 106; Conservative 59; Mismatches 159; Indels 41; Gaps 12;

QY 74 AEDERH-----FSECEPIPREAGEAYDFITSDMPPEQRFRLANQVGM 121  
 DB 36 AFVHRHADVLTVASDPGVYSSQLRPSQALSEQLISVIDPMMHRLRLVSAQATP 95  
 QY 122 PVYDELNRIQELACSLIESLRPGOC--NFTEDYAEPPPIRIFMLAGLPEEDIDPLKYL 180  
 DB 96 RIVADLERVETLAGQLLDVA--DGDTPVLVADFAVPLPVIVIAELGVPDPDRLLFRSM 153  
 QY 181 TDQMR-----PDGSMTFEAKKALYDYLPIIFQRQKPGTDAISIVA 224  
 DB 154 SDRMLQWVADPADMQFGDDADEDYORLYKEPRAMHAYLHDHVTDRRARANDLISALV 213  
 QY 225 NCQVNGRPTSEAKRMCGLLVGLDLYVNFLSMERLAKSPENRQLRPER--IP 282  
 DB 214 AARVGEERLTDEQIVEFGALLMAGHVSTMDLGNTVLCKDHP--RAEAAARADRSILIP 271  
 QY 283 AACCELLR-RFSLVADGRILTSDFEHGVQLKKDQIILPQMLSGIDERRKACPMHVD 340  
 DB 272 ALIEVLRRLRPITVMARKTTKDYVLAGTTIPAG--RMVPSLLSANHDAQVETDPDHL 330  
 QY 341 SRQVSHTEFGHSHLCLGHLARREIIVTLKEMLTRIPDFSIARGAQIQ-HKSGIVSGV 399  
 DB 331 AREG-RQIAFGHGHIYCGAPLARLEGRLALBALFRDFPSPDCAKIRYRDGLF-GV 388  
 QY 400 QALPL 404  
 DB 389 KNLPL 393

RESULT 9  
 R13350  
 ID R13350 standard; Protein: 403 AA.  
 AC R13350;  
 DT 05-JUN-1991 (first entry)  
 DE Cytochrome enzyme P450SU2.  
 KW Cytochrome P450; P450SU1; P450SU2; herbicide resistance.  
 OS Streptomyces griseolus.  
 PN W09103561-A.  
 PD 21-MAR-1991.  
 PR 27-AUG-1990; U04785.  
 PR 11-SEP-1989; US-405605.  
 PR 12-JAN-1990; US-464499.  
 PR 23-AUG-1990; US-569781.  
 PA (DUPO) DU PONT DE NEMOURS CO.  
 PI Dean C. Harder PA, Ieto KJ, Lichtner FT, Odell JT;  
 PI O'Keefe DP, Omer CA, Romesser JA;  
 DR WPI: 91-102077/14.  
 DR N-PSDB: Q11127.  
 PT DNA encoding cytochrome P450 enzymes - and electron donating  
 PT iron-sulphur proteins, used to confer herbicide resistance to  
 PT plants and microorganisms  
 PS Clam 15; page 158; 224pp; English.  
 CC This cytochrome P450 enzyme, P450SU2, is expressed alongside the iron  
 CC sulphur protein FeS-A, by a DNA sequence contained in a recombinant  
 CC plasmid. Host Streptomyces species are transformed with the plasmid  
 CC and are used to coat a plant seed to transform the plant. The res-  
 CC ultant transformants are resistant to herbicides.  
 CC See also Q11126.  
 SQ Sequence 403 AA;

Query Match 14.8%; Score 322; DB 1; Length 403;  
 Best Local Similarity 28.4%; Pred. No. 1.6e-24;

Matches 103; Conservative 64; Mismatches 166; Indels 30; Gaps 13;  
 QY 63 WIATGQLIREAYEDYRHFSSEC-----PIPREAGEAYDFIPT--SMDPPOROPALA 115  
 DB 52 WLVTNRQDVRAVLGDR-PSADAHRTGFFPLTAGREIIGTNPTRLRMDPEHARRML 110  
 QY 116 NOVGMPPVVDKLENRIQELACSLIESLR--QGOCNFTEDYAEPPPIRIFMLAGLPEEDI 174  
 DB 111 TADFIVKRVKAEARPEVQRIADOLVDRMTGTGRTSADLVTEFALPLSLVLCILLAGPYEDH 170  
 QY 175 PHL-----KULTQMRPDCSMTFEAKKALYDYLPIIFQRQKPGTDAIS-IVANGQVN 229  
 DB 171 AFQERSRVLTLRSTPE--EVRAQDELLEYLRLARTKRERDDAITSIRVARGEID 227  
 QY 230 GPRITSDEAKRMCGLLVGLDLYVNFLSMERLAKSPENHQLIERERIPACEELL 289  
 DB 228 DPQIAT-----MGRLLVARGELDDTOIATMGLLRNPDQARLARPAALVKGAVEELL 282  
 QY 290 RRFSLVADG--RILTSDFEHGVQLKKDQIILPQMLSGIDERRKNACP--MHVDSRQKV 345  
 DB 283 RULTIVHNVPRLATIEDVLIIGRTIAGSGVLC--MISSANNDAEVFPGGDDLDVARDAR 340  
 QY 346 SHTEFGHSHLCLGHLARREIIVTLKEMLTRIPDFSI-PCAQIQHKSGL-VSGVALP 403  
 DB 341 RHVAFGFGVHOCIGPLARVELQIAIETLLRLPLRLAVPHEEIPFGDMAIYGVHSLP 400  
 QY 404 LVM 406  
 DB 401 IAW 403

RESULT 10  
 R38309  
 ID R38309 standard; Protein: 412 AA.  
 AC R38309;  
 DT 04-DEC-1993 (first entry)  
 DE Sequence of the P450-soy protein.  
 KW P450soy; soy; gene; oxidation; haem protein.  
 OS Streptomyces griseus ATCC 13273.  
 PN W09312236-A.  
 PD 24-JUN-1993.  
 PR 16-DEC-1992; U10885.  
 PR 16-DEC-1991; US-807001.  
 PA (DUPO) DU PONT DE NEMOURS & CO E I.  
 PI Omer CA, Sarisiani FS, Trower MK;  
 DR WPI: 93-214178/26.  
 DR N-PSDB: Q45369.  
 PT Constitutive expression of P450 SOY (SOYC) and ferredoxin soy  
 PT (soyb) in Streptomyces - used for oxidn. of organic chemicals  
 PS Example: Figure 2; 45pp; English.  
 CC Cytochrome P450soy was purified from *S. griseus* ATCC 13273. Two  
 CC similar forms of P450soy were isolated. P450soy-delta, is derived  
 CC from P450soy by in vitro proteolysis during isolation. One of the  
 CC tryptic peptide fragments of cytochrome P450soy and of of the  
 CC P450soy-delta protein were subjected to automated degradation. The  
 CC NH2 terminal sequences are given in R38306 and R38307. A mixture of  
 CC oligos that consist of possible DNA sequences that could encode the  
 CC AAs FGVRQCL of the tryptic peptide was made. It consists of the  
 CC sequences in Q43290-23. The oligo mixture was end-labeled and used  
 CC to probe the EMBL4 library of *S. griseus* DNA. Hybridising plaques  
 CC were isolated and a 4.8kb SacI DNA fragment was isolated from one  
 CC clone that hybridised to the oligo probe mixture. As segment of the  
 CC 4.8kb fragment was sequenced and found to contain an ORF. Within  
 CC this ORF was a section that corresp. exactly to the AA sequence  
 CC determined from the cytochrome P450soy tryptic peptide (see Q45369,  
 CC R38309). The gene encoding the P450soy protein was called soyC. Five  
 CC nucleotides downstream from the stop codon for soyC another ORF was  
 CC identified. This ORF encodes an apparent ferredoxin-like protein.  
 CC The gene was designated soyB and the protein ferredoxin-soy.  
 SQ Sequence 412 AA;

Query Match 14.5%; Score 316; DB 1; Length 412;

Best Local Similarity 27.2%; Pred. No. 6.6e-24;  
Matches 84; Conservative 56; Mismatches 163; Indels 6; Gaps 5;

```
QY 103 MDPPROGFRALANOVMPVNDKLENIQIOLACSLIESLRPG-QCNFTEDYAEPPRIR 161
Db 105 VDDEPHNTQRMRLPTFSVKRIGALPRIOETVDRLLAMEQCGPAELVSFALPVSWM 164
QY 162 IFMLLAGLPEEDIPHLKYLIDQMTPRDGSMTFAEAKALYDYLPIPIOROKRPGTDAIS 221
Db 165 VICALLGVPAVDHAEFEERSQRLRPGADDVNNRARDLEELGALLIDRKRAEPEDGLD 224
QY 222 IVANGOVNGRPTSDCAKRMCGLLVGLDVTNVNFLSFSMEFLAKSPENROELIERPRI 281
Db 225 ELIHNDHPDGVDRQDLVAFVILLIAGHETTANNISLGFTLLSHRQDLARAGSTST 284
QY 282 PACEELLRRPSLVADG--RIITSDPEFHGVLKQDOILLPOMSLGDERKNACPMHYD 339
Db 285 AVVVEELL-RFLSTAEGIQRLATEMEVDGATIRKGEVVESTLINRDADVPPRAETLD 343
QY 340 FSRQKVSHTTEFGHSHLCLGOHLARREITVTLKEMLTFRIPDSIA-PSAOIQHKSQ-IVS 397
Db 344 WDRPARHHLACFGVHQCIGQNLARAEIDIAMRTLFRLLPGLRLAVPAHETIRHKRGDTIQ 403
QY 398 GVQALPLVW 406
Db 404 GLLDLPVAM 412
```

## RESULT 11

```
R51368
ID R51368 standard; Protein; 410 AA.
AC R51368:
DT 24-NOV-1994 (first entry)
DE Protein containing Cytochrome P450 SCA-2 activity.
KW Cytochrome P450 SCA-2; Streptomyces carbophilus;
OS Streptomyces carbophilus.
FH Key Location/Qualifiers
FT protein 1..410
FT /label= cytochrome P-450 containing activity
PN J06070780-A.
PD 15-MAR-1994.
PF 28-AUG-1992; 229969.
PR 28-AUG-1992; JP-229969.
PA (SANKY ) SANKYO CO LTD.
DR WPI: 94-128679/16.
DR N-PSDB: 061452.
PT Cytochrome P450 SCA-2 gene - from Streptomyces carbophilus,
PT useful in treatment of hyperlipidaemia
PS Claim 1; Page 13-14; 18pp; Japanese.
CC R51368 shows a protein having cytochrome P-450 activity. P-450
CC SCA-2 can be prepared commercially for use in the preparation of a
CC drug for the treatment of hyperlipidaemia.
SQ Sequence 410 AA;
```

Query Match 14.4%; Score 313.5; DB 1; Length 410;  
Best Local Similarity 25.6%; Pred. No. 1.2e-23;  
Matches 92; Conservative 63; Mismatches 185; Indels 19; Gaps 7;

```
QY 63 WIATRGQILREAYEDYR-----HFSSECPFIP--REAGEAYDFIPTSMDDPEOQOR 112
Db 56 WVVTKEHKAARLADPRLLSSDRHADFPATSPRFKAFQSGPAF----IGMDPEHGRIR 111
QY 113 ALANQOVGMFVNDKLENIQIOLACSLIESLRPG-QCNFTEDYAEPPRIRIFMLLAGPE 171
Db 112 RMTISEFTVKRIKQMRDVERIVGFIIDMLAAGPTADLVSOFAIPVPSWVICMILGVPY 171
QY 172 EDIHLVLTQOMTRPPDSMTFAEAKALYDYLPIPIEORQRQGTDAISIVANGOVNGR 231
Db 172 ADHEFPDASAKRLVQAVDADSAVARDFERYLDGLITKLESEFGTIGLGLVTHQDLADG 231
QY 232 PITSDEAKRMCGLLVGLDVTNVNFLSFSMEFLAKSPENROELIERPRIACEELLRR 291
```

Db 232 EIDRAELISTALLLVAGHETTASWTSLSVITLLENHPQDHAAIRADPSLVGAVEELLRV 291

QY 292 FSL--VADGRILTSYEPHGVQVKKGDQILLPOMSLGDERKNACPMHYDSRQKVSHTT 349

Db 292 LAIDADIAGRIATADIEIDGQILIRAGEGVIVTNSANDSSVFENPDRLDHRSAHHLS 351

QY 350 FGHSGLCLGOHLARREITVTLKEMLTFRIPDSIA-PSAOIQHKSQ-IVSGVQALPLVW 406  
Db 352 FGYGHQCLGQNLARAELEVLTLYLFDRIPLRLAVPBOQLTRPGTTIGOVNELPYTW 410

## RESULT 12

```
R14724
ID R14724 standard; Protein; 404 AA.
AC R14724:
DT 26-JAN-1992 (first entry)
DE 6-hydroxylase encoded by Eryf gene.
KW C-6 hydroxylation; erythromycin; 6-deoxyerythromycin; antibiotics;
KW Saccharopolyspora; cytochrome P450 monooxygenase; ss.
OS Saccharopolyspora erythraea.
PN W09116334-A.
PD 31-OCT-1991.
PF 16-APR-1991; 002600.
PR 18-APR-1990; US-510483.
PA (ABBO ) ABBOTT LABORATORIES.
PI Weber JM.
DR N-PSDB: Q14548.
DR N-PSDB: Q14548.
PT New 6-deoxyerythromycin derivs. - are antibiotics with increased
PT acid stability, produced by cultivation of saccharopolyspora.
PS Disclosure; Fig 3; 56pp; English.
CC The eryf gene encodes the 6-hydroxylase component of the cytochrome
CC P450 monooxygenase system responsible for the hydroxylation of 6-de-
CC oxyerythronolide B to erythronolide B. Interruption of this step
CC results in the formation of deoxyerythromycin A and new derivatives
CC useful as antibiotic which have better stability against acids that
CC the corresponding erythromycins. Interruption of the reaction can
CC be effected by an insertion into the eryf gene of a plasmid, gene
CC replacement or chemical or light-induced mutagenesis.
CC See also Q14549.
SQ Sequence 404 AA;
```

Query Match 13.6%; Score 296.5; DB 1; Length 404;  
Best Local Similarity 24.2%; Pred. No. 6.3e-22;  
Matches 90; Conservative 74; Mismatches 173; Indels 35; Gaps 10;

```
QY 63 WIATRGQILREAYEDYRHFS-----ECPTIPREAG--EAYDFIPTSM--DDPEQ 108
Db 40 WLVTGYDEAKAALSDLRSSDPKKKYPGEVEF-PAYLGFPEDEVNRYFATMTGSDPPTH 98
QY 109 ROFRALANQOVGMFVNDKLENIQIOLACSLIESLRPGQCNFTEDYAEPPRIRIFMLLAG 168
Db 99 TRLKRLVSOEFTVRRVEMRPREVOITAELDEVGDSGVVDIVYFAHPAPRIKVICELIG 158
QY 169 LPEEDIPHLKYLTD-----QMTPRDGSMTFAEAKALYDYLPIPIEOROKRPGTDAI 220
Db 159 VDF-----KYRGEGRMSSITLVMDPERAQOAAAREVNFITLDVERKRTPEGDLL 212
QY 221 SIVANGQ--VNGRPTSDCAKRMCGLLVGLDVTNVNFLSFSMEFLAKSPENROELIERP 278
Db 213 SALIRYQDDDDGGR-LSADELSSIALYVLLAGFEASVSLIGTGTYLLLTHRQDLAVRDR 271
QY 279 ERIPACEELLIRPSL-VADGRILTSYEPHGVQVKKGDQILLPOMSLGDERKNACPMH 337
Db 272 SALPNAAVEELIRYALQETTRFAAEDELICGVAIPQSYFLVANGAANRDPKQFPDR 331
QY 338 VDFSRQKVSHTTEFGHSHLCLGOHLARREITVTLKEMLTFRIPDSIARGAO--IQHSGI 395
Db 332 SDVITRGRHLSFGQGHFCMGRLAKLEGEVALRALGFRPALSLGIDADDVWRRSVL 391
QY 396 VSGVQALPLVWD 407
```



KW Daunomycin C-14 hydroxylase; doxa gene; doxorubicin; daunomycin;  
KW 13-dihydrodaunomycin; carminomycin; anthracycline;  
OS anticancer; cytostatic; cancer; therapy; plasmid pANT199.  
OS Chimeric - Streptomyces sp. strain C5.  
OS Chimeric - synthetic.  
FH Key location/Qualifiers  
FT Cleavage-site 31  
FT Protein /note= "enterokinase cleavage site"  
FT .474  
FT /note= "native daunomycin C-14 hydroxylase"  
PN W09744439-A2.  
PD 27-NOV-1997.  
PE 22-MAY-1997; U08690.  
PR 24-MAY-1996; US-653650.  
PA (OHIS ) UNIV OHIO STATE RES FOUND.  
PI Desanti CL, Dickens ML, Strohl WA;  
DR WPI: 98-018495/02.  
DR N-PSDB: V01452.  
PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase  
PT - also hydroxylation and oxidation of other anthracycline(s) with  
PT the same enzyme  
PS Disclosure; Fig 9; 59pp; English.  
CC This protein comprises an N-terminal modified enzyme derived from  
CC the daunomycin C-14 hydroxylase (see W36128) of Streptomyces sp.  
CC strain C5. It is encoded by a gene construct (see V01452) in  
CC plasmid pANT199. In this plasmid, the doxa gene (see V01447) is  
CC translationally fused with a leader sequence encoding 6  
CC histidine residues so that the fusion protein can be affinity  
CC purified on a nickel-agarose gel. Daunomycin C-14 hydroxylase  
CC is a P450-like enzyme capable of converting daunomycin to the  
CC anticancer agent doxorubicin. Host cells, especially Streptomyces  
CC host cells, transformed with pANT199 can be used in methods for  
CC the production of doxorubicin from daunomycin or for the  
CC hydroxylation and oxidation of other anthracyclines.  
SO Sequence 474 AA;

Query Match 7.38; Score 160; DB 1; Length 474;  
Best Local Similarity 23.68; Pred. No. 6.7e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 QEAMNVIQESNVPDLVTRCNGHIAIRGQILREAYEDYRHFSE--CPFIPREAGEAY 96  
DB 83 REAGPVV-EVNAP-----AGSPAWVITDDALAREVLADPRFKGPDLPAMRGVDDGL 135  
QY 97 D-----FIPTSMDPPEORFALANQVGMVVDKLENNIOELACSLFESL-----R 143  
DB 136 DIPVELRPFTLLIIVDGEDHRLKRIINAPNPRRLAERTDRIAAIADRLTEIADSSDR 195  
QY 144 POGQCNFTEDYAEPPPIRIFMLAGLP-----EEDIPHLKYL-----TDQMPRPD 188  
DB 196 SGEPAELLIGFAVYHNPILLVICELGVPTDPMAREAVGLKALGLGSPQASAGDGTDP 255  
QY 189 GSMTFEAKKEALYDYLIIIEQOROKPSTDALSIYANGVNGRPITSDPAKRMGCLLVG 248  
DB 256 GDVPDTSALFESL--LLEAVNAAKRDRTMTRVLYERAQAEFGSVSDQDLVMTTGLIFA 313  
QY 249 GIDTVVNFILSFSMEFLAKSPERQELTERPERIPACCELRLR-----FSLVADGRILTS 303  
DB 314 GHDTGGSFLGF--LLAEVLNAGRLAADADGALISFVEALRNHPRPVYSL--WRFPAAT 367  
QY 304 DYEFHGVQALKGDDILLFQMLSLGLDERKNACPMHNVDFSRQVSHITTFGSHLCLGQHIA 363  
DB 368 EVVIRGVRLPRGAPVLDIEGTNTDGRHNDAPNHAFNPRPSRRRLTFGDPHYVICIGBOLA 427  
QY 364 RREIIVTLKEWLTIRPDPSIA-PGAQIQ--HKSGLVSGVALPLVW 406  
DB 428 QLESRTMIGVLRSPQARLAIVPYEELRWCKRGAQTARLTDLR-VW 472



THIS PAGE BLANK (00118)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:38:03 : Search time 66.92 seconds  
(without alignments)  
103.693 Million cell updates/sec

Title: US-09-246-451-11

Perfect score: 2180

Sequence: 1 TTFETISNNALAPLPHVPE.....IVSGVQALPLVMPDATTKAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 16761186 residues  
Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 10%  
Listing first 45 summaries

Database :  
1: Issued\_Patents\_AA.\*  
2: /cgn1\_7/ptodata/1/1aa/5A.COMB.pep.\*  
3: /cgn1\_7/ptodata/1/1aa/5B.COMB.pep.\*  
4: /cgn1\_7/ptodata/1/1aa/6.COMB.pep.\*  
5: /cgn1\_7/ptodata/1/1aa/6CTUS.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	365.5	16.8	416	3	US-09-320-878-18
2	340	15.6	403	5	5212296-9
3	330	15.1	406	5	5212296-6
4	316	14.5	412	1	US-08-102-863-11
5	316	14.5	412	4	PCT-US92-10885-11
6	271.5	12.5	419	3	US-09-335-409-8
7	160	7.3	422	2	US-09-096-982-5
8	160	7.3	422	2	US-08-653-650A-5
9	160	7.3	422	2	US-09-096-982-8
10	160	7.3	422	2	US-08-653-650A-8
11	157	7.2	443	2	US-09-096-982-9
12	157	7.2	443	2	US-08-653-650A-9
13	155	7.1	422	1	US-08-396-218-2
14	155	7.1	422	1	US-08-760-116-2
15	130	6.0	512	2	US-08-194-981E-5
16	123.5	5.7	382	3	US-09-320-878-7
17	118	5.4	513	3	US-08-948-564-6
18	113.5	5.2	490	1	US-08-201-118-3
19	113.5	5.2	490	1	US-08-201-118-9
20	113.5	5.2	490	2	US-08-238-821B-9
21	113.5	5.2	490	2	US-08-238-821B-9
22	113.5	5.2	490	4	PCT-US95-05744-3
23	113.5	5.2	490	4	PCT-US95-05744-9
24	113	5.2	501	3	US-08-906-791-2
25	111.5	5.1	490	1	US-08-201-118-7
26	111.5	5.1	490	2	US-08-238-821B-7
27	111.5	5.1	490	4	PCT-US95-05744-7
28	111.5	5.1	504	1	US-08-457-274A-25

29	111.5	5.1	504	4	PCT-US95-05758-25	Sequence 25, Appl
30	110	5.0	492	3	US-08-724-466B-2	Sequence 2, Appl
31	105.5	4.8	496	1	US-08-313-075A-50	Sequence 50, Appl
32	104	4.8	497	3	US-08-724-466B-4	Sequence 4, Appl
33	102.5	4.7	472	2	US-08-622-166A-2	Sequence 2, Appl
34	102.5	4.7	472	2	US-08-622-166A-4	Sequence 4, Appl
35	99.5	4.6	490	1	US-08-201-118-1	Sequence 1, Appl
36	99.5	4.6	490	2	US-08-238-821B-1	Sequence 1, Appl
37	99.5	4.6	490	4	PCT-US95-05744-1	Sequence 1, Appl
38	99.5	4.6	498	1	US-08-457-274A-24	Sequence 24, Appl
39	99.5	4.6	498	4	PCT-US95-05758-24	Sequence 24, Appl
40	96.5	4.4	500	2	US-08-314-601-2	Sequence 2, Appl
41	96.5	4.4	500	4	PCT-US95-13051-2	Sequence 2, Appl
42	94.5	4.3	523	3	US-08-606-505B-67	Sequence 67, Appl
43	94.5	4.3	1026	1	US-07-998-003A-95	Sequence 95, Appl
44	94.5	4.3	1026	1	US-08-453-274B-95	Sequence 95, Appl
45	94.5	4.3	1026	1	US-08-453-695A-95	Sequence 95, Appl

ALIGNMENTS

RESULT 1  
US-09-320-878-18  
: Sequence 18, Application US/09320878A  
: Patent No. 6117659  
: GENERAL INFORMATION:  
: APPLICANT: ASHLEY, Gary  
: APPLICANT: BETLACH, Melanie C.  
: APPLICANT: BETLACH, Mary C.  
: APPLICANT: MCDANIEL, Robert  
: APPLICANT: TANG, Li  
: TITLE OF INVENTION: RECOMBINANT NABONOLIDE POLYKETIDE SYNTHASE  
: FILE REFERENCE: 300622002120  
: CURRENT FILING DATE: US/09/320, 878A  
: EARLIER FILING DATE: 1999-05-27  
: EARLIER APPLICATION NUMBER: CIP OF 09/141,908  
: EARLIER FILING DATE: 1998-08-28  
: EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
: EARLIER FILING DATE: 1998-05-06  
: EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
: EARLIER FILING DATE: 1997-04-30  
: EARLIER APPLICATION NUMBER: 60/119,139  
: EARLIER FILING DATE: 1999-02-08  
: EARLIER APPLICATION NUMBER: 60/100,880  
: EARLIER FILING DATE: 1998-09-22  
: EARLIER APPLICATION NUMBER: 60/087,080  
: EARLIER FILING DATE: 1998-05-28  
: NUMBER OF SEQ ID NOS: 34  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 18  
: LENGTH: 416  
: TYPE: PRT  
: ORGANISM: Streptomyces venezuelae  
US-09-320-878-18

Query Match 16.8%, Score 365.5; DB 3; Length 416;  
Best Local Similarity 28.3%, Pred. No. 8e-30;  
Matches 97; Conservative 62; Mismatches 163; Indels 21; Gaps 7;

OY	76	EDYRHFSECEPIPEAGCAVDIFTSMDPEBQRFALANQVGMVYDKLENRIQELA	135
DB	72	KDWR--NSTPLTEAEALNHNMLESS--DPRHRTLRKLVAREFTMRVRELLRPVQELIV	127
OY	136	CSLIESSL--RPGGOCNFEDEVAEPPRIRIFMLGLGTPREDIPHLKYLTDQMTTRPGSMTF	193
DB	128	DGLVDAMLAAPDGRADLMESLAWPLITVISLGLVPPDPDAAFVFWTDAFVFPDDPAQA	187
OY	194	AEAKALYDLVLIPIEORRKGPTDAIS--IVANGVNGRPITSDEAKRMCCILLVGIGDT	252
DB	188	QTANMMSGYISRLIDSRKGGDGEDLSALVRTSDEDSRLTSEELGMHILLVAGHET	247

[illegible]

Query Match	15.6%;	Score 340;	DB 5;	Length 403;
Best Local Similarity	28.9%;	Pred. No. 3.5e-27;		
Matches	105;	Conservative	63;	Mismatches 165;
		Indels	30;	Gaps 13;

  

QY	63	WIATRGQILREAYEDYRHFSSSEC----	PFIPRAGEAYDFIPT--	SMDPEQRQFRALA	115
Db	52	WLVTIRHODYRAVLGDPR-FSADAHRTGPFLLAGCREILITGTPFLRMDDPEHARLRML			110
QY	116	NOQVGMVPVVDLENNIOELACSLIESLRP-QGQCFTFDYAPFPFIRIFMLAGPREDI			174
Db	111	TADFLVKKVEVMREPVOYRADDLDVRMTTGKRSAGLVTEFALPLPSVLICLLGVYEDH			170
QY	175	PHL-----KYLTDDMTRPDGSMTFAEAKBALDYDLPIIEQRQKPGTAAIS-IVANGOVN			229
Db	171	AFQFQRSHVLLTLRSTPE--EVRAAOELLEYLRLARTKRERDDAIIISLTVARGELD			227
QY	230	GRPTISDEAKMKGGLLVGGLDYVNFISFSKSEFLAKSPENHOELIERPERIPAAACELL			289
Db	228	DTOIAT-----MGRLLIYAGHETTANMMLSTLVLLRNPDOLARLRABALVKKAGVEELL			282
QY	290	RRESLIVADG--RIILSDVEFHVOLKKGDDIILPOMLSGLDERKNACP--MHVDSROKV			345
Db	283	RYLTIVNHGVRIRATEDVILIGRTITAAEGYVC--MISSANDAEVFGGDDLDVARDAR			340
QY	346	SHITFGHSGHCLGQHLARREIIVTLKEMLRIRIPDESIA-PGAQIQHKSGI-VSGVQALP			403
Db	341	RHVAFGFGVHOCLGQPLARVELQIATIELRLRLRLRAVLAVPHEEIPFGDMAITGVHSIP			400
QY	404	LVW 406			
Db	401	IAW 403			

  

RESULT	3
5212296-6	
Patent No. 5212296	
APPLICANT: DEAN CAROLINE HARDER, PATRICIA A. LETO, KENNETH J. O'KEEFE, DANIEL P. OMER, CHARLES A. ROMSSER, JAMES A.	

```

TEPPERMAN, JAMES M.
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
; CYTOCHROMES
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/569,781
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 464,499
; FILING DATE: 12-JAN-1990
; APPLICATION NUMBER: 405,605
; FILING DATE: 11-SEP-1989
; SEQ ID NO:6:
; LENGTH: 406
; 5212296-6

```

```

QY 31 PSNLSAGV--EAAVAIOESVPLWTRCGG-ATATGOLIREAYEYR-----79
Dd 17 PSNSCEQLPDGAYQLRDTGPRLHRTYLIDGRAMAVYTHAARAKLLGDPRLSSNFTDD 76
QY 80 HESSECPFP--REAGEAYDEIPTSMDPPEROFALANOVGMPVVDKJENRIQELACS 137
Dd 77 NFAPTSPEFAEVRSPQAF----IGDPPEHGTRRRMTISEFYKRIKGRNPEVEEYVHG 132
QY 138 LIESLRIOG--OCNTEDYAEFPFPRITMLLAGPEEDIPILKYTLTDMTRPDGSMYFAEA 196
Dd 133 FLIDEMLAAGPTADLVOSFALPVPBMVICRLGVADYADHEFFOASKRIVOSTAQSALITA 192
QY 197 KEALYDYLPIIEORCKPGDAI-SIVANOVGGRITSDKARKMGKLLVGGDLVYN 255
Dd 193 RNDLAGYIDGLITFOEPREGALVGAIVAOQLANGE-IDREELI STAMLLIAGHETTAS 251
QY 256 FLFSMEFLAKSPENROELIERPERIPAAECBELLRFSL--VADGRILTSDYEPHGVLK 313
Dd 252 MTSLSVITLIDHPRQYALRADRSRLVGYGABELLRYLAIADIAGRAVATADIEYEGILIR 311
QY 314 KGDOITLPLQMLSGIDERRKMACPMHVDERSOKVSHTTFGHSGSHLCLGONLARREIYTLKE 373
Dd 312 AGEGVIVANSIANDGTVYEDPDALDIHRSARHHLAGFGVHOCIGONLARLEVEYLINA 371
QY 374 WLTRIPDESLA-PCAQIQHKSG-IYVSQVALPLVW 406
Dd 372 LMDRVPTLRILAVPVEOLVLRGTTIGGVNELPLVW 406

```

RESULT 4  
 US-08-102-863-11  
 : Sequence 11, Application US/08102863  
 : Patent No. 546590  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: SAKISLANI, SIMA  
 : TITLE OF INVENTION: CONSTITUTIVE  
 : TITLE OF INVENTION: EXPRESSION OF P45  
 : TITLE OF INVENTION: AND FERREDOXIN-SO  
 : TITLE OF INVENTION: STREPTOMYCES  
 : NUMBER OF SEQUENCES: 12  
 :  
 : CORRESPONDENCE ADDRESS:  
 :  
 : ADDRESSEE: E. I. DU PONT DE NEMOURS  
 :  
 : ADDRESSEE: E. I. DU PONT  
 : STREET: 1007 MARKET STREET  
 : CITY: WILMINGTON  
 : STATE: DELAWARE  
 : COUNTRY: USA  
 : zip: 19898  
 :  
 : COMPUTER READABLE FORM:  
 :  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS



SOFTWARE: PatentIn Release #1.0,  
SOFTWARE: Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/102,863  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/807,001  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CALLEGOS, R. THOMAS  
REGISTRATION NUMBER: 32,692  
REFERENCE/DOCKET NUMBER: CR-9000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-892-7342  
TELEFAX: 302-892-7949  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 412 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-102-863-11

Query Match 14.5%; Score 316; DB 1; Length 412;  
Best Local Similarity 27.2%; Pred. No. 1.2e-24;

Matches 84; Conservative 56; Mismatches 163; Indels 6; Gaps 5;

QY 103 MDPPEQRFALANQVGNPVVDKLENRIOLACSLIESLRPOG-QCNFTEDYAEPPPIR 161  
DB 105 VDDPEHNTQRRMLITPFSVKRIGALRPRIQETVDRLLDMERQGPRAELVSFAFLPVSM 164  
QY 162 IFMLLAGPREDIPHLKYLTDQMPDPSMTFAEAKALYDYLPIIEOROKPGTDAIS 221  
DB 165 VICALGVYADNAHAFEEESQRLRGPGADVNRARDELEEVYLGALIDRKRAEPDGLD 224  
QY 222 IVANGVNGRPITSDAKRMCGLLVGGIDTVVNFLSFMEFLASPEHROELIERPERI 281  
DB 225 ELIHDHDPGPDYDRQOLVAFVILLIAGHETPANNISLGFTLLSHPEQLALRAGGTST 284  
QY 284 PACCELLARFESLVADG--RLITSDYEFHGVOLKKGDQILLPQMSGDERKNACPMHYD 339  
DB 285 AVVVEELL-REFLSIAEGQRLATEDEMDVGATIRKGEVFTSLINDADVFPRAETLD 343  
QY 340 FSRQKVSHTTFCGSHLCLGQHLARREIIVTLKEMLTIRIPDSIA-PGAQIQHKGK-IVS 397  
DB 344 WDRPARHHLAFEGVHOCIGQNLARAEIDIAMRTLFEKRLPGRLAVPAHEIRHKPGDTIQ 403  
QY 398 GVQALPLVW 406  
DB 404 GLDLPLVAM 412

RESULT 5  
PCT-US92-10885-11

; Sequence 11, Application PC/TUS9210885  
; GENERAL INFORMATION:  
; APPLICANT: SARIASLANI, SIMA  
; TITLE OF INVENTION: CONSTITUTIVE  
; TITLE OF INVENTION: EXPRESSION OF P450SOY  
; TITLE OF INVENTION: AND FERREDOXIN-SOY IN  
; TITLE OF INVENTION: STREPTOMICES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: USA  
; ZIP: 19898

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch,  
MEDIUM TYPE: 1.0 MB  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh System, 6.0  
SOFTWARE: Microsoft Word, 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/10885  
FILING DATE: 19921216  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: CALLEGOS, R. THOMAS  
REGISTRATION NUMBER: 32,692  
REFERENCE/DOCKET NUMBER: CR-9000-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-892-7342  
TELEFAX: 302-892-7949  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 412 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
PCT-US92-10885-11

Query Match 14.5%; Score 316; DB 4; Length 412;  
Best Local Similarity 27.2%; Pred. No. 1.2e-24;

Matches 84; Conservative 56; Mismatches 163; Indels 6; Gaps 5;

QY 103 MDPPEQRFALANQVGNPVVDKLENRIOLACSLIESLRPOG-QCNFTEDYAEPPPIR 161  
DB 105 VDDPEHNTQRRMLITPFSVKRIGALRPRIQETVDRLLDMERQGPRAELVSFAFLPVSM 164  
QY 162 IFMLLAGPREDIPHLKYLTDQMPDPSMTFAEAKALYDYLPIIEOROKPGTDAIS 221  
DB 165 VICALGVYADNAHAFEEESQRLRGPGADVNRARDELEEVYLGALIDRKRAEPDGLD 224  
QY 222 IVANGVNGRPITSDAKRMCGLLVGGIDTVVNFLSFMEFLASPEHROELIERPERI 281  
DB 225 ELIHDHDPGPDYDRQOLVAFVILLIAGHETPANNISLGFTLLSHPEQLALRAGGTST 284  
QY 284 PACCELLARFESLVADG--RLITSDYEFHGVOLKKGDQILLPQMSGDERKNACPMHYD 339  
DB 285 AVVVEELL-REFLSIAEGQRLATEDEMDVGATIRKGEVFTSLINDADVFPRAETLD 343  
QY 340 FSRQKVSHTTFCGSHLCLGQHLARREIIVTLKEMLTIRIPDSIA-PGAQIQHKGK-IVS 397  
DB 344 WDRPARHHLAFEGVHOCIGQNLARAEIDIAMRTLFEKRLPGRLAVPAHEIRHKPGDTIQ 403  
QY 398 GVQALPLVW 406  
DB 404 GLDLPLVAM 412

RESULT 6  
US-09-335-409-8

; Sequence 8, Application US/09335409  
; Patent No. 6121029  
; GENERAL INFORMATION:  
; APPLICANT: Schnupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goetlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/335,409  
; CURRENT FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 8  
LENGTH: 419  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-335-409-8

Query Match 12.5%; Score 271.5; DB 3; Length 419;  
Best Local Similarity 23.8%; Pred. No. 5,3e-20;  
Matches 100; Conservative 67; Mismatches 179; Indels 75; Gaps 14;

OY 6 OSNANLALPRVPHVLPDPMNPSNLSAGVDF---ANAVLOESVNPDLVWTRCNG 61  
DB 3 QEQANQSESTKP-----AFDEKPFAP-----GYAEDPPAIEKLEEA--TPIFYWD--BGR 48  
OY 62 HMIATRGQ-----LIREAYEDYRHFSSECPPIPREAGAVDFIPTSMPPQR 109  
DB 49 SWVLTRYHDVSAVFDEDFPANSREEMESSAETSSAIP---ELSDMKRYGLFGLPEPDHA 104  
OY 110 QFRALANQVGMPPVDKLENRIQELACSLIESLRPGOCNFTEDYAEPPPIRIFMLAGL 169  
DB 105 RFRKLVNPSFTSRALDLRAELQRTVDQLDARSQEEFEDVROVAGEIPMAISALLKV 164  
OY 170 PEEIDPLKLYLTDQMTPRDGSWTFPAEAEALYDILP-----I 207  
DB 165 PAE-----CDEKFRFRFGSAT---ARALGVGLVPQVDEETKTLVASVTEGLALLHDV 212  
OY 208 TEORROKP-GTDAISIVANGVNGRPITSDEAKRMCGLLVGLDTVNNLSFSMEFLAK 266  
DB 213 LDERRRNLENDVLTMLQAEADGSRSLTKELVALVGAILAGDTTTYLLAFVAILLR 272  
OY 267 SPEHROELIERPERIPACEELLRRFSLVADG--RILTSDEYFNGVOLKKGDOI--LLPO 322  
DB 273 SPEALELVKAERGLARNALDEVLRFDNILLRIGTVAFARQDLEYCASIKKGMVLLIPS 332  
OY 333 MUGSDEKKNACPMHVDPSRKQVSHTTGSHLCLGQHLARREIIVLKEKLEIIPDS 382  
DB 333 ALR--DGTVFSPDVEDVRRDTGASLAVGRPHVCPSLARLEAIVGTIFRRPEMK 390  
OY 383 I 383  
DB 391 L 391

RESULT 7  
US-09-096-982-5  
Sequence 5, Application US/09096982  
Patent No. 5962293  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEE, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/096,982  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Goltick, Mary E.  
REGISTRATION NUMBER: 34829

REFERENCE/DOCKET NUMBER: 22727/00131  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 216-622-8458  
TELEFAX: 216-241-0816  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-096-982-5

Query Match 7.3%; Score 160; DB 2; Length 422;  
Best Local Similarity 23.6%; Pred. No. 2,3e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

OY 39 QEAVNAVLOESVNPDLVWTRCNGMIATRGOLIREAYEDYRHFSSE--CPPIPREAGEAY 96  
DB 31 REAGPVV--EVNAP-----ACGPAMVITDDALAREVLADPRFVKGPDLAPAMRGVDDGL 83  
OY 97 D-----FPTSMDPPEQOFALANQVGMVVDKLEKRIQELACSLIESL-----R 143  
DB 84 DIPVELRPFTLIVADGDRRLRIHAPAFNPRLAERTDRIADRLTTELADSSDR 143  
OY 144 PGOCNFTEDYAEPPPIRIFMLAGLP-----EEDIPHLKYV-----TDQMTRPD 188  
DB 144 SGEPALLEGFAHYHFPPLVITCELLGVPTDPMAREAVGVALKALGPGSAGDGTDP 203  
OY 189 GSNMFAEAKREALYDILPIEORROKPTDAISIVANGVNGRPITSDEAKRMCGLLV 248  
DB 204 GDVDTGSALIESL--LLEAVNAARRKDFRTMTVLYERAQAEFGSVSDQLYVMTGLIFA 261  
OY 249 GLDTVNNLSFSMEFLAKSPRHROELIERPERIPACEELLR-----FSVYAGRLITS 303  
DB 262 GHDTTGSFLGF--LLAVYLAGRLAADADGALISRYVEEALRHHPVYSI--WRPAT 315  
OY 304 DYFRNGVOLKKGDOIILPQMLSGLDERKNACPMHVDPSRKQVSHTTGSHLCLGQHLA 363  
DB 316 EVVIRGVNLRPGARVYLDIEGTNTDGRHNDAPHFHPRDRSRRRLITDGDHYIGQGLA 375  
OY 364 RREIIVLKEMLTRIPEFSIA--PGAQIO--HKSGLVSGVALPIVW 406  
DB 376 QLESRTMIGVLSRFPQARLAVPYEELRMCKKGAQTARLTDLR--VW 420

RESULT 8  
US-08-653-650A-5  
Sequence 5, Application US/08653650A  
Patent No. 5976830  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEE, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/653,650A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

```

: NAME: Gollick, Mary E.
: REGISTRATION NUMBER: 34829
: REFERENCE/DOCKET NUMBER: 22727/00131
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 216-622-8458
: TELEFAX: 216-241-0816
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 422 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-653-650A-5
```

Query Match 7.3%; Score 160; DB 2; Length 422;

Best Local Similarity 23.6%; Pred. No. 2.3e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

```

QY 39 QEAMAVLQESNVPLDWMTRCNGHWIATRGQILREAYEDYRHFSSE--CPFIPEAGGAY 96
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 31 REAGPVV-EVNAP-----AGGPAAVITDDALAREVLADPRFVKGPDLAFTAMRGVDGL 83
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 97 D-----FIPTSMDPEORQFALANQVGMVYDKLENRIQELACLISLESL-----R 143
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 84 DIPVELRPFTLIAVDGSDHRLRRIHAPFNPRLAERTDIRIAIADRLLTLELADSSDR 143
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 144 POGOCNFTEDYAEPRPIKIFMLAGLP-----EEDIPHLKYL-----TDQMTRPD 188
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 144 SGEPAELIGGFAHYHPRLLVICELGVPTDPAAREAVGLKALGSGPQSGAGGCTPRA 203
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 189 GSMFPAEAKELAYDLPIIEORROKPGTDALISIVANGVNGRPITSDPAKRMGCLLVG 248
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 204 GDVPRTSALLESLL--LLEAVHAAKKDPTMTRVLYERAQAEGSVSDQLVYMIGLIFA 261
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 249 GLDTVYNELFSMELTASPEHROELLERPERIPACCELLR-----FSLVADRIILTS 303
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 262 GHDPTGSFLGF--LLAEVLAGRLAADADGALISFVEEALRNHNPVYSL--WRFAT 315
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 304 DYEFHGVOLKKGDQILLPOMLSGLDERKNACRMHNVDSRQVSHTFPGSHILGQHILA 363
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 316 EVVIRGVALLPRGAPVLDIEGTNTDGRHNDAPNHNPRPSRRLLTFDGRHYICGEOLA 375
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 364 RREIIVTLKEMLTIRIPDSIA-PCGAQIO--HKSGIVSGVALPLVW 406
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 376 QLESRTMIGVLRSPQARLAVPYEELRMCKRGAQTARTLDLP-VW 420
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
```

```

RESULT 9
US-09-096-982-8
: Sequence 8, Application US/09096982
: Patent No. 5962293
: GENERAL INFORMATION:
: APPLICANT: Strohl, William R.
: APPLICANT: Dickens, Michael L.
: APPLICANT: Desanti, Charles L.
: TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CALFEE, HALTER & GRISWOLD
: STREET: 800 Superior Avenue, Suite 1400
: CITY: Cleveland
: STATE: Ohio
: COUNTRY: USA
: ZIP: 44114-2688
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: FILING DATE: US/09/096,982
```

```

: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Gollick, Mary E.
: REGISTRATION NUMBER: 34829
: REFERENCE/DOCKET NUMBER: 22727/00131
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 216-622-8458
: TELEFAX: 216-241-0816
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 474 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-096-982-8.
```

Query Match 7.3%; Score 160; DB 2; Length 474;

Best Local Similarity 23.6%; Pred. No. 2.8e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

```

QY 39 QEAMAVLQESNVPLDWMTRCNGHWIATRGQILREAYEDYRHFSSE--CPFIPEAGGAY 96
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 83 REAGPVV-EVNAP-----AGGPAAVITDDALAREVLADPRFVKGPDLAFTAMRGVDGL 135
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 97 D-----FIPTSMDPEORQFALANQVGMVYDKLENRIQELACLISLESL-----R 143
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 136 DIPVELRPFTLIAVDGSDHRLRRIHAPFNPRLAERTDIRIAIADRLLTLELADSSDR 195
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 144 POGOCNFTEDYAEPRPIKIFMLAGLP-----EEDIPHLKYL-----TDQMTRPD 188
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 196 SGEPAELIGGFAHYHPRLLVICELGVPTDPAAREAVGLKALGSGPQSGAGGCTPRA 255
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 189 GSMFPAEAKELAYDLPIIEORROKPGTDALISIVANGVNGRPITSDPAKRMGCLLVG 248
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 256 GDVPRTSALLESLL--LLEAVHAAKKDPTMTRVLYERAQAEGSVSDQLVYMIGLIFA 313
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 249 GLDTVYNELFSMELTASPEHROELLERPERIPACCELLR-----FSLVADRIILTS 303
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 314 GHDPTGSFLGF--LLAEVLAGRLAADADGALISFVEEALRNHNPVYSL--WRFAT 367
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 304 DYEFHGVOLKKGDQILLPOMLSGLDERKNACRMHNVDSRQVSHTFPGSHILGQHILA 363
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 368 EVVIRGVALLPRGAPVLDIEGTNTDGRHNDAPNHNPRPSRRLLTFDGRHYICGEOLA 427
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 364 RREIIVTLKEMLTIRIPDSIA-PCGAQIO--HKSGIVSGVALPLVW 406
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 428 QLESRTMIGVLRSPQARLAVPYEELRMCKRGAQTARTLDLP-VW 472
: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
```

```

RESULT 10
US-08-653-650A-8
: Sequence 8, Application US/08653650A
: Patent No. 5976830
: GENERAL INFORMATION:
: APPLICANT: Strohl, William R.
: APPLICANT: Dickens, Michael L.
: APPLICANT: Desanti, Charles L.
: TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CALFEE, HALTER & GRISWOLD
: STREET: 800 Superior Avenue, Suite 1400
: CITY: Cleveland
: STATE: Ohio
: COUNTRY: USA
: ZIP: 44114-2688
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
```



MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/653,650A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gollick, Mary E.  
REGISTRATION NUMBER: 34829  
REFERENCE/DOCKET NUMBER: 22727/00131  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 216-622-8458  
TELEFAX: 216-241-0816  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 443 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-653-650A-9

Query Match 7.2%, Score 157; DB 2; Length 443;  
Best Local Similarity 23.6%, Pred. No. 5,2e-08;  
Matches 95; Conservative 50; Mismatches 210; Indels 48; Gaps 13;

QY 39 QEAMAVLQESNVPDLWTRCNGHMIATRGOLIREAYEDYRHSSE--CPITPREAGEAY 96  
DB 52 REAGYV--EVNAP-----ACGPAAVITDDALAREVLADPREVKRPDLAPAMRGVDGL 104  
QY 97 D-----FIPTSMDPPEQOFRLANOVGMRYVDKLEKIOELASLIESL-----R 143  
DB 105 DIPVELRPFTLLIADGSDHRRRLRIHAPAFNPRLAERTORIAIAIDLITELADSSDR 164  
QY 144 POGOCNFTEDYAEPRIRIFMLAGLP-----EEDIPHLKYL-----TDQMTRPD 188  
DB 155 SGERALLEGFAVHFRPLVICELGVPYTDPMAREAVGLKALGPGSGSGGTTPA 224  
QY 189 GSMFAEAKKALYDYLRPIIEQRORPGTDAISIVANGVNGRPITSDANKMCGLLVG 248  
DB 225 GDVPTSALESF--LLEAVHAAKRRKDTRTMTRVLYERAQAEGSVSDQLVYMTGLIFA 282  
QY 249 GLDPTVNVLSMFERLAKSPENROELIERPERIPACCELLRFSLY--ADGRILTSYE 306  
DB 283 GHDTTGSFLGF--LLAEVLAGRLAADADGALISRFVEEALKNRPVYTYLMRFATEV 339  
QY 307 FHGVOLKKGQDILLPQMLSGDERKNACPMHVDFSROKVSHTTFGSHLCLGHLARE 366  
DB 340 IRGVALPRGAVLVLDIEGTNTDGRHNDARHAFNRPDRSRRLITFGDGRHYICGDLADE 399  
QY 367 IIVTLKEMLTIRPDFSIA--PGAQIO--HKSGIVSGVQALPLVW 406  
DB 400 SRTMIGVLRSPQARLAVPYEELRWCKRGAQTARLTDLR--VW 441

RESULT 13  
US-08-396-218-2  
; Sequence 2, Application US/08396218  
; Patent No. 5695966  
; GENERAL INFORMATION:  
; APPLICANT: INVENTI, Augusto  
; APPLICANT: BREME, Umberto  
; APPLICANT: COLOMBO, Anna L  
; APPLICANT: HUTCHINSON, Charles R  
; APPLICANT: OTTEN, Sharee  
; APPLICANT: SCOTTI, Claudio  
; TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIKAIIDO, MARWELSTEIN, MURRAY & ORAM

STREET: 655 Fifteenth Street, N. W., Suite 330 - G  
STREET: Street Lobby  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/396,218  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: KITTS, Monica C  
REGISTRATION NUMBER: 36,105  
REFERENCE/DOCKET NUMBER: P1615-5002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/638-5000  
TELEFAX: 202/638-4810  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-396-218-2

Query Match 7.1%, Score 155; DB 1; Length 422;  
Best Local Similarity 23.8%, Pred. No. 7.8e-08;  
Matches 97; Conservative 51; Mismatches 202; Indels 58; Gaps 15;

QY 39 QEAMAVLQESNVPDLWTRCNGHMIATRGOLIREAYEDYRHSSECPITPREAGEAYDF 98  
DB 31 REAGYV--EVNAP-----ACGPAAVITDDALAREVLADPREVKD-----PDLAPAAARG 78  
QY 99 IPTSM--PREQOFRLA-----NOVGMRYVD-----KLEKIOELASLIESL-- 142  
DB 79 VDDGIDIPVELRPFTLLIADGSAHRRRLRIHAPAFNPRLAERTORIAIAGRLITELA 138  
QY 143 ---RPOGOCNFTEDYAEPRIRIFMLAGLP-----EEDIPHLKYL-----TDQ 183  
DB 139 DASGRSGKPAELIGFAVHFRPLVICELGVPYTDPMAREAVGLKALGPGSGGSD 198  
QY 184 MTRPDGSMFAEAKKALYDYLRPIIEQRORPGTDAISIVANGVNGRPITSDANKMCG 243  
DB 199 GTDPAAGVDPDTSALBSL--LLEAVHSARNDPTMTRVLYERAQAEGSVSDQLVYMIT 256  
QY 244 LILVGLDPTVNVLSFSMFERLAKSPENROELIERPERIPACCELLRFSLY--ADGRIL 301  
DB 257 GLIFAGHDTTGSFLGF--LLAEVLAGRLAADADDAVSRFVEEALKNRPVYTYLMRFA 313  
QY 302 TSDVEFHGVOLKKGQDILLPQMLSGDERKNACPMHVDFSROKVSHTTFGSHLCLGHL 361  
DB 314 ATEVTIGVLRPRGAVLVLDIEGTNTDGRHNDARHAFNRPDRSRRLITFGDGRHYICG 373  
QY 362 LARREIIVTLKEMLTIRPDFSIA--PGAQIO--HKSGIVSGVQALPLVW 406  
DB 374 LAQLESRTMIGVLRSPQARLAVPYEELRWCKRGAQTARLTDLR--VW 420

RESULT 14  
US-08-760-116-2  
; Sequence 2, Application US/08760116  
; Patent No. 5786190  
; GENERAL INFORMATION:  
; APPLICANT: INVENTI, Augusto  
; APPLICANT: BREME, Umberto  
; APPLICANT: COLOMBO, Anna L  
; APPLICANT: HUTCHINSON, Charles R

APPLICANT: OTTEN, Sharee  
APPLICANT: SCOTTI, Claudio  
TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIKAIIDO, MARCELSTEIN, MURRAY & ORAM LLP  
STREET: 655 Fifteenth Street, N. W., Suite 330 - G  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentlin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/760,116  
FILING DATE: 3-DEC-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/396,218  
FILING DATE: 27-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: KITTS, Monica C  
REGISTRATION NUMBER: 36,105  
REFERENCE/DOCKET NUMBER: P1615-6007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/638-5000  
TELEFAX: 202/638-4810  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-760-116-2

Query Match 7.1%; Score 155; DB 1; Length 422;  
Best Local Similarity 23.8%; Pred. No. 7.8e-08;  
Matches 97; Conservative 51; Mismatches 202; Indels 58; Gaps 15;  
DB 31 REAGPVV-EVNAP-----AGSPAVITDOLAREVLADPRFVAD-----PDLPAAWRG 78  
QY 99 IPIYSMD--PPROQFALA-----NOVGMFVVD-----KLENRIQELACSLIESL- 142  
DB 79 VDDGLDIPVPELRPFLLIADVGEAHRRLRIHAPFNPRLAERTDRIAAIAGRLITELA 138  
QY 143 -----RPGQCFETEDYAEPPRIREFMLAGP-----EEDIDHLYL-----TDQ 183  
DB 139 DASGRSKPAELLIGGFVYHPELVICELGVPTDPMARAVSVLALGIGPOSGGD 198  
QY 184 MTRPDGSMTEFAEAKALYDYLPIIEQROKPGTDAISIVANGOVNRPITISDEAKKMG 243  
DB 199 GTPDAGVPRDSALESL--LLENVHSARNDPTMTVILVERAAEGSGVSDDDLVYMIT 256  
QY 244 LLVVGGLDYVNNPLSFSMEPLAKSPHEHROELIERPERIPACACELLRRFSLV--ADGRIL 301  
DB 257 GLTFAGHDITGSEGLF--LLAEVLAGRILADADEDAVSREVEBALRYHHPVPTLWRA 313  
QY 302 TSDVEFHGVOLKKGQDILLQMLSGLDERKNACPMHNDFSQKVSHTTFGGSHLCLGQH 361  
DB 314 ATEVTTIGVRLPRGAPVLDVIEGNTGCRHNDADHAFHPDRPSMRRLTFGGGPHYCIGEQ 373  
QY 362 LARREIIVTLKEMVLRIPDESIA-PGAQIO--HKSGIVSGVQALPLVW 406  
DB 374 LAQLESTMTIGVLRSPREARLAVPYDELKRCRKGQATARLTLP-VW 420

RESULT 15  
US-08-194-981E-5  
Sequence 5, Application US/08194981E  
Patent No. 5886157  
GENERAL INFORMATION:  
APPLICANT: GUENGERICH, F. Peter  
APPLICANT: GUO, Zuyu  
APPLICANT: SANDHU, Punam  
APPLICANT: GILLAM, Elizabeth M. J.  
TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF  
CYTOCHROME P450  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NEEDLE & ROSENBERG, P. C.  
STREET: Suite 1200, 127 Peachtree Street, NE  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303-1811  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentlin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/194,981E  
FILING DATE: February 10, 1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Elizabeth Selby  
REGISTRATION NUMBER: 38,298  
REFERENCE/DOCKET NUMBER: 22000.0022  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 688-0770  
TELEFAX: (404) 688-9880  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 512 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
FRAGMENT TYPE: N-terminal  
US-08-194-981E-5

Query Match 6.0%; Score 130; DB 2; Length 512;  
Best Local Similarity 22.4%; Pred. No. 4.3e-05;  
Matches 95; Conservative 67; Mismatches 158; Indels 104; Gaps 22;  
DB 107 DLYTFLLISNGQSMFS-----PD-----SGPVAAAR-----RLAQGLKSFSTASD 149  
QY 82 -----SSQCFPIREAGEAYDFIPTSMD-----PPEOROPR-----ALANOVGMFVVDKLEN 130  
DB 150 PASSTCYLEHNSKEAELVISTIOELMAGRGHNPYRYVAVSTVNICICGRRIDHN 209  
QY 131 IOELACSLIESLRPQCGCFETEDYAEPPRIREFMLAGLPEEDIDPHLYLTDQMTRPDGS 190  
DB 210 HOEL-LSLV-----NLNNNGE-----VSGSNPAEFILRYL-----PNPS 246  
QY 191 M-TFAEAKALYDYLPIIEQRO-----KQETDAISIVANGQ-----VNGRPITSDE 237  
DB 247 LNAFKDINEKEYSFMQKVEKHKTFPEKGHIRDTTD--SLIEHQEKQLDENANVOLSD 304  
QY 238 AKRRCGLLVG-GIDTVVNNLSFSMEPLAKSPHEHROELIE-----RP-----ERI 281  
DB 305 KIINIYVLDLFGAGFDVTTAISWSLMTLVNMPVYORIOEELDTVIGRSSRPLSDRSHL 364  
QY 282 P---ACCEELLRRFSLV--ADGRITSDVEFHGVOLKKGQDILLQMLSGLDERKNACP- 345

Db	365	PYMAFILETFRHSSEVPETIPHSSTRDTSJKGFYIPKRCVFVNOMQINHDOKLWNP	424
OY	336	-----MHVDFSRKV---SHTTFGSHLCLGOHLARREIIVTLKEMLTRIPDPSIAP	385
Db	425	EFLPERFLPDGAIKVLSEKVIIFGKGRKCIGETVARWEVFLFLAILLQRV-EFSVPL	483
OY	386	GAOI	389
Db	484	GVKV	487

Search completed: October 4, 2000, 12:38:05  
Job time: 21422 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:39:34 : Search time 79.83 Seconds  
(Without alignments)  
359.568 Million cell updates/sec

Title: US-09-246-451-11  
Perfect score: 2180  
Sequence: 1 TTEIQTQNNALPLPPHVE.....IVSGVQALPLVMDPATTKAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL\_12:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_podent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	365.5	16.8	416	2	087605 streptomyc
2	364	16.7	396	2	052544 amycolatops
3	364	16.7	397	2	059523 micromonos
4	347	15.9	410	2	09X5P9 streptomyc
5	346.5	15.9	404	2	054302 streptomyc
6	344.5	15.8	407	2	059819 streptomyc
7	329	15.1	411	2	032460 actinomyc
8	326	15.0	406	2	059723 pseudomon
9	325.5	14.9	420	2	092H01 streptomyc
10	324.5	14.9	376	2	031785 bacillus su
11	321	14.7	407	2	09X5P8 streptomyc
12	318	14.6	310	2	032927 mycobacteri
13	317	14.5	406	2	087192 streptomyc
14	316	14.5	410	2	09X5P7 streptomyc
15	313.5	14.4	410	2	059831 streptomyc
16	312	14.3	400	2	09XDB0 mycobacteri
17	303.5	13.9	397	2	024727 nocardioide
18	303	13.9	417	2	059910 streptomyc
19	301	13.8	406	2	085697 streptomyc

20	294.5	13.5	337	2	052816 amycolatops
21	293	13.4	351	2	052572 amycolatops
22	291.5	13.4	406	2	087675 amycolatops
23	291	13.3	388	2	P77977 streptomyc
24	279	12.8	421	2	052561 amycolatops
25	277.5	12.7	310	2	P96562 amycolatops
26	267	12.2	398	2	087674 amycolatops
27	265.5	12.2	437	2	092FC0 mycobacteri
28	264	12.1	403	2	052560 amycolatops
29	262.5	12.0	411	2	09X803 streptomyc
30	261.5	12.0	386	2	059921 streptomyc
31	261	12.0	395	2	09X5P2 streptomyc
32	256	11.7	426	2	092FC3 mycobacteri
33	255	11.7	406	2	052823 amycolatops
34	250	11.5	411	2	060005 putative
35	248.5	11.4	511	2	052569 amycolatops
36	239	11.0	433	2	09XCG6 streptomyc
37	226	10.4	391	2	052822 amycolatops
38	225	10.3	398	2	059079 amycolata a
39	224.5	10.3	391	2	087673 amycolatops
40	215	9.9	397	2	052802 amycolatops
41	203.5	9.3	419	2	050242 agrobacteri
42	182.5	8.4	119	2	085655 streptomyc
43	179	8.2	120	2	085653 streptococ
44	174.5	8.0	313	2	09X418 myxococcus
45	164	7.5	120	2	085650 streptomyc

## ALIGNMENTS

### RESULT 1

ID 087605 PRELIMINARY: PRT: 416 AA.

AC 087605;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE CYTOCHROME P450 MONOOXYGENASE.  
GN PICK OR PICK.  
OS Streptomyces violaceus (Streptomyces venezuelae), and  
OS Streptomyces venezuelae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC15439;  
RA BETLACH M.C., KEALEY J.T., ASHLEY G.W., MCDANIEL R.;  
RT "Characterization of the macrolide P450 hydroxylase from Streptomyces  
RT venezuelae which converts narboxycin to plicomycin.",  
RL Biochemistry 0:0-0(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.VENEZUELAE; STRAIN=ATCC15439;  
RA MEDLINE: 99051447.  
RT XUE Y., WILSON D., ZHAO L., LIU H.-W., SHERMAN D.H.;  
RT "Hydroxylation of macrolactones YC-17 and narboxycin is mediated by  
RT the PICK-encoded cytochrome P450 in Streptomyces venezuelae.",  
RL Chem. Biol. 5:661-669(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.VENEZUELAE; STRAIN=ATCC15439;  
RA MEDLINE: 98445333.  
RT XUE Y., ZHAO L., LIU H.-W., SHERMAN D.H.;  
RT "A gene cluster for macrolide antibiotic biosynthesis in Streptomyces  
RT venezuelae: architecture of metabolic diversity.",  
RL Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).  
DR EMBL: AF087022; AAC64105.1; -;  
DR EMBL: AF079139; AAC68886.1; -;  
DR HSSP: Q00441; LOXA.  
DR PFAM: PF00067; P450; 1.  
DR PRINTS: PR00359; BP450.  
KW Monooxygenase.



QY	224	ANGGVNNGRPITSDEKRRKCGILLVGGIDTVVNFLSFMSNEFLAKSPENHOELIEREERIPA	283
Db	210	VOARQODSLSEQLELDLAIIGLVAGVSTTTQIADFPVYLLMTRELRQLDLDRPELLPS	269
QY	284	ACEELLRRFSL---VADGRILTSDEYEFHGVOYLKKQDQILLPOMLSGLDERKNACPMHDF	340
Db	270	AVEELTRVNPVIGVGAFPRVAVEDVTLLGVTIIRAGEPVLTAIGCANNRQAOFRDADRIDV	329
QY	341	SRQKVSHTTFEGHSLCLGCHLARREIIVTLKEWTRIPDSIA-PCAQIQHKSQ-IVSQ	398
Db	330	DRTPNQHGFHGHVHCIGAPLARVEILOVALEVLRLRPIRIGIPETQLRWSEGMLLRG	389
QY	399	VOALPLVW	406
Db	390	PLELPVW	397
RESULT	4		
Q9X5P9			
ID	Q9X5P9	PRELIMINARY;	PRT; 410 AA.
AC	Q9X5P9		
DT	01-NOV-1999	(Tremblrel. 12, Created)	
DT	01-NOV-1999	(Tremblrel. 12, Last sequence update)	
DT	01-NOV-1999	(Tremblrel. 12, Last annotation update)	
DE	CYTCHROME P450 HYDROXYLASE ORF3.		
OS	Streptomyces lavendulae.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-NRRL 2564.		
RX	MEDLINE: 99201491.		
RA	MAO Y.Q., VAROGU M., SHERMAN D.H.;		
RT	"Molecular characterization and analysis of the biosynthetic gene		
RT	cluster for the antimitor antibiotic milomycin C from Streptomyces		
RT	lavendulae NRRL 2564."		
RL	Chem. Biol. 6:251-263(1999).		
DR	EMBL: AF127374; AAD28449.1; -		
SO	SEQUENCE 410 AA; 45887 MW; 507C7F38 CRC32;		
Query Match	15.9%; Score 347; DB 2; Length 410;		
Best Local Similarity	27.0%; Pred. No. 2.2e-20;		
Matches 113; Conservative 69; Mismatches 193; Indels 44; Gaps			14;
QY	9	ANLAPLRP-HVPEHLVPEDMKVNDPSNLSAGVOEAMVLDQESVDPDLVWTRCGNHMIATR	67
Db	15	AGEBAPFPHAPDRDRL-EPDPY-----WEPLRRERPLQRLVTLPGGEAMLATR	60
QY	68	GOLAREAVEDXYRHNSSE---CRPIPREAGEAVDFP-----TSMPPRQORORALAN	116
Db	61	YODRAVAFAD-RRSRQDLAVABGAR-----FLPHORPPDAVLVSQGPDAIRRLRLV	112
QY	117	QVAGMPVVDKLENRIQELACSLIESLRQEG-OCNTEDEYAEFPFRIRIMLAGLPEEDIP	175
Db	113	KVLPFRAREDEMRPLQIRFADGLIDAMEEMGRPADLVEDFSLPFAVSMICELLYGVERDRK	172
QY	176	HLKILTLDMTPRDSMTFPAEKE---ALYDULPIITIEDROKPGTDAISIVANGOVNRP	232
Db	173	RCFVMSDALILTTT-AHPRAQVDDYMMQNHDIRLIGLVAAQRVRYPTADLLIGSLVTABDEBK	231
QY	233	ITTSBEAKMCGCLLVVGGIDTVVNFLSFMSMERLAKSPENHOELIEREERIPAACEELLRRF	292
Db	232	LTEGELVRLAAILIAGTETSASQIPNFIYVLRHQQLLEKIRINDHDLIPDAVEELRFV	291
QY	293	SL-VADG---RLTSDYEFHGVOYLKKQDQILLPOMLSGLDERKNACPMHDFPSRKVSHTT	349
Db	292	PICGVADGPRPATEDVELGVLVAGERIVVPSMGANRDPDLFTPDDELIDLARPNPHLG	351
QY	350	FGHSHSLCLGCHLARREIIVTLKEWTRIPDSIA-PCAQIQHKSQ-VSGVOALPLVW	406
Db	352	FGAPRHHICGLAARVELQITLTLFRRYPRRLAIVPEESLSWKSGLWVRGNHTPVM	410

RESULT	5			
ID	054302	PRELIMINARY:	PRT:	404 AA.
AC	Q54302;			
DJ	01-NOV-1996 (TREMBLER)	01,	Created)	
DJ	01-NOV-1996 (TREMBLER)	01,	Last sequence update)	
DJ	01-NOV-1999 (TREMBLER)	12,	Last annotation update)	
DE	CYTOCHROME P450.			
GN	RABN.			
OS	Streptomyces hygroscopicus.			
OC	Bacteria; Filicutes; Actinobacteria; Actinobacteridae;			
CC	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-NRRL 5491.			
RX	MEDLINE; 95372374.			
RA	SCHWECKE T., APARICIO J.F., MOLNAR I., KOENIG A., KHAM L.E.,			
RA	HAYDOCK S.F., OLITYNYK M., CAFFEREY P., CORTES J., LESTER J.B.,			
RA	BOHM G.A., STAUNTON J., LEADLAY P.F.;			
RT	"The biosynthetic gene cluster for the polyketide immunosuppressant			
RT	ravanycin.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 92:7839-7843(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-NRRL 5491.			
RA	APARICIO J.F., MOLNAR I., SCHWECKE T., KOENIG A., HAYDOCK S.F.,			
RA	KOENIG A., STAUNTON J., LEADLAY P.F., LESTER J.B., BOHM G.A.,			
RL	Gene 0:0-0(0).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-NRRL 5491.			
RA	APARICIO J.F., MOLNAR I., SCHWECKE T., KOENIG A., HAYDOCK S.F.,			
RA	EH KHAM L., STAUNTON J., LEADLAY P.F., LESTER J.B., BOHM G.A.,			
RA	STAUNTON J., LEADLAY P.F.;			
RL	Gene 0:0-0(0).			
DR	EMBL; X86780; CAA60465.1; -.			
DR	HSSP; Q00441; IOXA.			
DR	PFAM; PF00067; P450. 1.			
DR	PRINTS; PR00359; BP450.			
SO	SEQUENCE 404 AA; 45071 MW; 05AB94DF CRC32;			

Query Match 15.9%; Score 346.5; DB 2; Length 404;  
 Best local Similarity 28.6%; Pred. No. 2,4e+20;  
 Matches 102; Conservative 63; Mismatches 177; Indels 15; Gaps 9;

```

        63 WIAIRGQLIREAYEYRIHFSECF--IPREAGANYF-IPTSMDDPEQGFRLANOV 118
           ||| : : : : | | | | | : : : : : ||||| : : : :
       Db 50 WLASMEVVAVFVDPR-FRSATLGKVPRALPIQQPYIMLMDDPEHTRLRVAVTKA 108
           ||| : : : : | | | | | : : : : : ||||| : : : :
       QY 119 VGMPEVDAIKENRIELACLISLESIRPOG-QCNFTEDYAEPPIRIFMLLAGIPEPDIPH 177
           : : : : | | | | | : : : : : ||| : : : : | : : : |
       Db 109 LTSRMELARPTEGEVADDLLDKMAKAPADMEDFALLPIRNICELLGVPIEDQKF 168
           : : : : | | | | | : : : : : ||| : : : : | : : : |
       QY 178 KYLDQMTRPGSGMT--FAEKALUYLIPIIQOROKRESTAISIVANGVGRPT 234
           : : : : | | | | | : : : : : ||| : : : : | : : : |
       Db 169 RTWSQOML-SNGAYSQEYVMAAGSLYLULSELIAERRKKOPTNDLSLVLRARDKDRLS 227
           : : : : | | | | | : : : : : ||| : : : : | : : : |
       QY 235 SDEAKRMGGLLYGGDLTVNVAFLEFSMEFLKSPENHOELIERPERIPACEELLRRPSL 294
           | : : : | : : : : | : : : : | : : : : | : : : |
       Db 228 ETELVGFAVTLTIACYETTAIGAVALIQRHSANTDPKYVDHPREIDFIHRTSNPMSIG 287
           : : : : | | | | | : : : : : ||| : : : : | : : : |
       QY 295 VADG---RLTSDVEFHGVLKKGDQILLPMLSGLDERKNACPMHYDVSNOQVSHTFEG 351
           | : : : | : : : : | : : : : | : : : : | : : : |
       Db 288 AKQASWVMAVEDELSETGIKAGEAVALIQHNHSANTDPKYVDHPREIDFIHRTSNPMSIG 347
           : : : : | | | | | : : : : : ||| : : : : | : : : |
       QY 352 HGSHLICGONHLARRPIITLKEMILFRIDPFLA-PGAOIQKSG-IYGVQAQLPLVW 406
           ||| : : : : | : : : : | : : : : | : : : : | : : : |
       Db 348 HGHHCMGAOLVRVKMOTALLSLIRPALRPAPERIKFLRGTLVPSELEPLTW 404
    
```

```
RESULT 6
OC 059819 PRELIMINARY; PRT; 407 AA.
ID 059819
AC 059819:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE CYTOCHROME P450 (EC 1.14.14.1).
GN OLEP.
OS Streptomyces antibioticus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95255619.
RA RODRIGUEZ A.M., OLANO C., MENDEZ C., HUTCHINSON C.R., SALAS J.A.;
RT "A cytochrome P450-like gene possibly involved in oleandomycin
biosynthesis by Streptomyces antibioticus.";
RL FEMS Microbiol. Lett. 127:117-120(1995).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: L37200; AAA92553.1; -.
HSSP: 000441; 10XA.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
DR PRAM: PF00067; P450; 1.
DR PRINTS: PR00359; BP450.
KM OXidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT BINDING 356 356 HEME (BY SIMILARITY).
SQ SEQUENCE 407 AA; 44957 MW; 04411C60 CRC32;
```

```
Query Match 15.8%; Score 344.5; DB 2; Length 407;
Best Local Similarity 28.1%; Pred. No. 3.5e-20;
Matches 104; Conservative 63; Mismatches 168; Indels 35; Gaps 11;
```

```
QY 60 GGHMIATRGQLREAVEDYRHFSSEC---PTPREAGEAVDFPT-----SMDPEQ 108
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 50 GTAWLVTRMSDARIVLGDSDR-FSTAATDPATPR-----MPTPEPDGVLADPPDH 101
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 109 RQFRLANQVYGMVVDKLENRIQELACSLIESLRPOG-OCNTEDEYAEPPPIRIFMLA 167
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 102 TRLRLVYGFARVREEMPRVSLVSDLDVANGSPADVEFLAVPPVAVTELL 161
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 168 GLPEEDIPHLKYTLTDQ---TRPDGSMTEFAEKALYDYL--PIEORRQKGTDAIS 221
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 162 GVPLEDRDLRTFESDAMLSSTR---LTAAEIQRVQODFMVYMGVLVAQRDADTEDLG 217
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 222 IVANGQVNGRPITSDKAKKRCGLLVGGLDTVYVNFSLFSMEFLAKSEHROELIERERI 281
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 218 ALALATNDNDHLTKGETVNNGVSLILAGHETSVNQITNLVLLTTERKRYESLVADPALV 277
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 PAACEELLRRFSVLAVDG---RLTSDYEFHGVOLKKGDQILLPOMLSGLDERKNACPMHV 338
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 278 PAAVEEMLRTPPLVSAGSFVAVATEDELSTVYRAGEPCPVVHNASNRDEEVDHDEL 337
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 339 DFSROKVSHTTFGHSILCLGOHLARREIIVTLKEMLTRIPDSIA-PGAQIQHKSG-IV 396
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 338 DFHHRNPHIAFGHGHCIGAGIQLGRLQDALVRRPPTLDLAPVAGLKKWQMLI 397
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 397 SGVQALPLVW 406
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 398 RGLERQIVSW 407
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 7
OC 032460 PRELIMINARY; PRT; 411 AA.
ID 032460
AC 032460:
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE ORF 10.
OS Actinomadura hibisca.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
```

```
OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;
OC Actinomadura.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=P157-2;
RC MEDLINE: 97480928.
RA DAIRI T., HAMANO Y., IGARASHI Y., FURUMAI T., OKI T.;
RT "Cloning and nucleotide sequence of the putative polyketide synthase
genes for pradiacin biosynthesis from Actinomadura hibisca.";
RL Biosci. Biotechnol. Biochem. 61:1445-1453(1997).
DR EMBL: D87924; BAA23153.1; -.
HSSP: 000441; 10XA.
DR PRAM: PF00067; P450; 1.
DR PRINTS: PR00359; BP450.
SQ SEQUENCE 411 AA; 44860 MW; 98A628B6 CRC32;
```

```
Query Match 15.1%; Score 329; DB 2; Length 411;
Best Local Similarity 26.4%; Pred. No. 6.4e-19;
Matches 114; Conservative 61; Mismatches 199; Indels 58; Gaps 16;
```

```
QY 4 TIGSNANLALPRPHVPEHLVDFEDMNP-----SNLSAGVDAAVLOESNVPLWTRCN 59
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 9 TVDPRPDVTPAFPRRPD-----DPQPRCEHAKLRASDPAKVVLP-----T 50
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 GGH-MIATRGQLREAVEDYRHFSSEC---PTPREAGEAVDFPT-----MDPE 107
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 51 GDHNVVTRYADRVFVSD-RRFSKEAVTRGADR-----LIPMGSKSLVIMDPE 102
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 QROFRLANQVGMVVDKLENRIQELACSLIESLRPOG-OCNTEDEYAEPPPIRIFMLA 166
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 103 HTRKRLIVSRATFARVREEMPRVSLVSDLDVANGSPADVEFLAVPPVAVTELL 162
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 167 AGLPEDIPHLKYTLTDQMTPDGSMTEFAEKE-----ALYDILPIIEORRQKGTDAI 220
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 163 LGVPEDPRPRQDMDTML-TIGAPALQADEIKAAVGRILGYLAELIDAKTAPADDL 221
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 221 SIYANGQVNGRPITSDKAKKRCGLLVGGLDTVYVNFSLFSMEFLAKSEHROELIERERI 280
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 222 SLSRAIADP-GLSEELLTFGMTLAAGYHTTAITHSYHLIRPSTRARLRREPSG 280
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 281 IPACEELLRRFSVLAVDG---RLTSDYEFHGVOLKKGDQILLPOMLSGLDERKNACPM 336
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 281 IPAAVEELL-RYGOIGGAGAIKRIAVDEVGGLVAGAVIPLFMAANRDPVFADPE 339
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 337 HVDPSROKVSHTTFGHSILCLGOHLARREIIVTLKEMLTRIPDSIA-PGAQIQHKSGI 395
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 340 ELDGRTDNDPHIALGHGIIHVCAGAPLARLEQVLLFTVVERTPALRLAIDADITWRPGL 399
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 396 V-SGVQALPLVW 406
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 400 AFARPDALPIAW 411
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 8
OC 059723 PRELIMINARY; PRT; 406 AA.
ID 059723
AC 059723:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE CYTOCHROME P450 LIN (EC 1.14.14.1).
GN LINC.
OS Pseudomonas incognita.
OC Bacteria; Proteobacteria.
RN [1]
RP SEQUENCE FROM N.A.
RA ROP J.D., GUNSAIUS I.C., SLIGAR S.G.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: L23310; AAA25810.1; -.
HSSP: P3306; 1CPT.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
```





Db 240 PDEPDPSPSSRHIAFAVSGHFCGALALRLAEATVTLTAISARPPQOLA-GEIYKPNV 298  
QY 395 IVSGVALPL 404  
Db 299 AMRGMSALPV 308

RESULT 13  
087192 PRELIMINARY; PRT; 406 AA.

ID 087192  
AC 087192; Q59915;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE ORF-2 PROTEIN.  
OS Streptomyces griseus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA UEDA K.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 95378086.  
RA UEDA K., KIM K.M., BEPPU T., HORINOCHI S.;  
RT "Overexpression of a gene cluster encoding a chalcone synthase-like  
RT protein confers reddish-brown pigment production in Streptomyces griseus.";  
RL J. Antibiot. 48:638-646(1995).  
DR EMBL: AB018074; BAA33494.1; -  
DR HSSP: Q00441; 10XA.  
DR PRAM: PF00067; P450.1.  
DR PRINTS: PR00359; BP450.  
SQ SEQUENCE 406 AA; 45739 MW; 950450FE CRC32;

Query Match 14.5%; Score 317; DB 2; Length 406;  
Best Local Similarity 27.7%; Pred. No. 5.9e-18;  
Matches 106; Conservative 47; Mismatches 202; Indels 28; Gaps 7;

QY 44 VLOBSNPDLVWTRNGCHWIAATRGOLIREAYEDYRHFSSCEPIPREGEAYDPIPTS- 102  
Db 32 LITREPYSRIMAYGEGEAMLVTR-----YEDVRYVTDTDRF-SASVALGRGFPRTPT 83  
QY 103 -----MDPEOROFALANOVVGMVYVKLENRIOLACSLIESLRPGQ-CNF 150  
Db 84 EPIVQAESINLMDPPASSRLGLVAKSTPRRVQMRGQTORVVDRLIDEMEEGSPDGF 143  
QY 151 TEDYAEFPPIRIFMLAGLPREDIPHLK--YLTQDMTRPDGSMFPAEAKKALYDYLPIIT 208  
Db 144 VARVASALPLITTTICELADIPADRPWLRAHAMTMNNVGAAGKQDAVRAKALRGTFOELT 203  
QY 209 EORROKPGTDAISIVANGOVNGRPITSDEAKRMCGLLVGLDITVNVFLSPFSMEFLAKSP 268  
Db 204 ADDRSPGEDLISLTARADODELLDDDELAVMAMVLLITGQDTTTLQGMIAVTLTLTRP 263  
QY 269 EHRBELIERPRIPAAECBELLRFSL---VADGRILTSDFYFHGVOLKKGQIILPOMLS 325  
Db 264 DLLSLAEAPRLRPTLEELRLHLPFRKGVGIPRIALEDVLSGLKAGDVVHVSYLTA 323  
QY 326 GLDERKNACPMHVPFSROKVSHTTEGSHLCLGSHLCLARRRLIYTLKMWLIRIP--DPSI 383  
Db 324 NRDSAKEDRPDELDPDRPTIPHTFTGCAHHCLGAPLATMELEVAFTLLTRFPALRLDV 383  
QY 384 APGAQIOHKSGIVSGVALPLW 406  
Db 384 PPEVSWMTSTIMWRYPLALPYTW 406

RESULT 14  
Q9X9P7 PRELIMINARY; PRT; 410 AA.

AC Q9X9P7;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE NIKF PROTEIN.  
GN NIKF.  
OS Streptomyces tendae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TUE901;  
RA BRUNNER C., LAUER B., SCHWARZ W., MOERLE V., BORMANN C.;  
RT "Molecular characterization of co-transcribed genes from Streptomyces  
RT tendae Tug901 involved in the biosynthesis of the peptidyl moiety of  
RT the peptidyl nucleoside antibiotic nikkomycin.";  
RL Mol. Gen. Genet. 0:0-0(0).  
DR EMBL: Y18574; CAB46536.1; -  
SQ SEQUENCE 410 AA; 45884 MW; 9B52CB74 CRC32;

Query Match 14.5%; Score 316; DB 2; Length 410;  
Best Local Similarity 27.1%; Pred. No. 7.2e-18;  
Matches 98; Conservative 64; Mismatches 180; Indels 20; Gaps 9;

QY 63 WIATRGOLIREAYEDYRHFSSCE--PTIP-----REAGEAYDPIPT-----SMDPPEQR 109  
Db 47 WLVLKHDARKLADPRVSADRLHPAPGRILTAQRITTEVRRLSTRSMIHLGDGDEHG 106  
QY 110 QFRALANOVGMVVDKLENRIOLA-CSLIESLRPGQCNFTEDYAEFPPIRIFMLAG 168  
Db 107 AHRKILGSEFSLRLIALRPVQELVDRSIDEMLAAPRALVHVSQAVNSLVICELLG 166  
QY 169 LPREDIPHL-KYLTQDMTRPDGSMFPAEAKALYDYLPIIT-EORROKPGTDAISIVANG 226  
Db 167 VPHEQRDRFHEMAGLVRSVSIRERAASDALNDFLEDITVEKERGEPDTDLIGRLIAR 226  
QY 227 QVNRPIITSDEAKRMCGLLVGLDITVNVFLSPFSMEFLAKSPERQELIERPEITPACE 286  
Db 227 NRRTPVMTIHDEIVGTAVMLTAGHQTANMTSLGVALLLENPEHAKRIARADPSLLPAYE 286  
QY 287 ELIRRFSLV--ADGRILTSDFYFHGVOLKKGQIILPOMLSGLDERKNACPMHVPFSROK 344  
Db 287 EMLTFYSVENAPARVATEDIEIGVTIRKDEGIVSSGLADNMDEVENHRDRIDFEGCA 346  
QY 345 VSHRTFGHSLCLGQILARRIITVTLKEMLTIRIPDSIA-PGAQIOHKSGI-VSGVAL 402  
Db 347 RHHVAFGVGHQICGOMLARVEIIVETTLIRVPGLSLAVAPABELPYKDDAGIYIVRV 406  
QY 403 PL 404  
Db 407 PV 408

RESULT 15  
059831 PRELIMINARY; PRT; 410 AA.

ID 059831  
AC 059831;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE CYCLOCHROME P450 SCA-2 (EC 1.14.14.1).  
OS Streptomyces carbophilus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SANK 62585;  
RX MEDLINE: 96001248.  
RA WATANABE T., NARA F., SERIZAWA N.;  
RT "Cloning, characterization and expression of the gene encoding  
RT cytochrome P-450sca-2 from Streptomyces carbophilus involved in  
RT production of pravastatin, a specific HMG-CoA reductase inhibitor.";





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:35:14 ; Search time 76.18 Seconds

(Without alignments)  
128.722 Million cell updates/sec

Title: US-09-246-451-12

Perfect score: 2179

Sequence: 1 TTETIQSNANLAPLPPHVE.....IVSGVOALPLVWDPATRKAV 414

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues  
Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: A.Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2126	97.6	405	Y04128	Pseudomonas putida
2	1178	54.1	446	Y04126	Bacterial and mamm
3	370	17.0	336	R60777	Myricanin IV hyd
4	338	15.5	408	R7867	S. clavuligerus Or
5	336.5	15.4	587	W33274	S. fradiae tylosin
6	336	15.4	406	R11349	Cytochrome enzyme
7	328	15.1	403	R11350	Cytochrome enzyme
8	327	15.0	411	R11389	Actinomadura hibi
9	323.5	14.8	398	W54389	Streptomycetes pr
10	319.5	14.7	410	W11585	Protein containing
11	315	14.5	412	R38309	Sequence of the p4
12	296.5	13.6	404	R14724	6-hydroxylase enco
13	196	9.0	398	R47521	Vitamin D hydroxyl
14	160	7.3	422	W36128	Daunomycin C-14 hy
15	160	7.3	474	W36132	N-terminal modifie
16	155	7.1	422	W00729	Daunorubicin 14-hy
17	148	6.8	533	R15057	Cytochrome P450C25
18	144	6.6	518	W67616	A. nidulans phenyl
19	143	6.6	494	R62825	Human steroid-21-h
20	135	6.2	512	W93216	Human cytochrome P
21	133	6.1	508	W35711	Chrysanthum flavon
22	133	6.1	512	R72365	Human auxillary cy
23	133	6.1	512	R93172	Human cytochrome P
24	132	6.1	512	R72365	Human auxillary cy
25	132	6.1	512	W00652	Cytochrome P450IA1
26	132	6.1	512	R93173	Human cytochrome P
27	131.5	6.0	524	R76544	Mitochondrial cyto
28	131	6.0	524	R76544	Rat hepato-cytochr
29	128.5	5.9	523	R59291	Rat liver cytochro
30	128.5	5.9	898	P61030	Entire coded sequ
31	128.5	5.9	898	P61082	Complete translati
32	128.5	5.9	899	P61056	translation of pla
33	128.5	5.9	1144	P81334	Expression prod. o

34	128.5	5.9	1150	1	P81335	Expression prod. o
35	128.5	5.9	1150	1	P81337	Expression prod. o
36	128.5	5.9	1162	1	P81336	Expression prod. o
37	127	5.8	512	1	R72364	Human auxillary cy
38	127	5.8	512	1	R93171	Human cytochrome P
39	125	5.7	519	1	R34881	Chimeric cytochrom
40	124	5.7	516	1	W67617	P. chrysogenum phe
41	122	5.6	507	1	W92994	Murine 1-alpha-hyd
42	122	5.6	507	1	W93491	Mouse 1-alpha-OHase
43	120.5	5.5	516	1	R72360	Human cytochrome P
44	120.5	5.5	516	1	W00183	Cytochrome P450IA2
45	120.5	5.5	516	1	R93167	Human cytochrome P

## ALIGNMENTS

RESULT 1

ID Y04128 standard; Protein; 405 AA.

DE 11-JUN-1999 (first entry)

KW Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;

KM oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;

OS Pseudomonas putida.

PN W09908812-A1.

PD 25-FEB-1999.

PE 17-AUG-1998; U16979.

PR 20-AUG-1997; US-056754.

PA (UYRP ) UNIV ROCHESTER.

PI Jones JP, Shimoji M;

DR WPI: 99-190131/16.

DR N-PSDB: X19926.

PT New P450 fusion proteins - comprising a portion of a bacterial

PT cytochrome P450 protein and a portion of a mammalian cytochrome P450

PT protein

PS Disclosure: Page 12-13; 51pp; English.

CC The present invention describes a fusion proteins comprising a portion

CC of a bacterial cytochrome P450 protein and also a portion of a mammalian

CC cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or

CC any compound having a carbon-hydrogen bond. The fusion protein can be

CC used for hydroxylating a compound to be oxidised. It can also be used in

CC the bioremediation of an environmental pollutant. Since the fusion

CC protein is soluble, it can be subject to structural elucidation by X-ray

CC crystallography for designing functional proteins. It can be readily

CC expressed in soil bacteria to facilitate bioremediation. The present

CC sequence represents Pseudomonas putida cytochrome P450 protein P450-cam

CC from the present invention.

SQ Sequence 405 AA:

Query Match 97.6%; Score 2126; DB 1; Length 405;  
Best Local Similarity 99.5%; Pred. No. 2.6e-212;  
Matches 403; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	10	NLAPLPHVPEHVLFDMDMNPSTLSGVOEAMVLDSENPVDLVWTRCGHIAIRGQ	69
DB	1	NLAPLPHVPEHVLFDMDMNPSTLSGVOEAMVLDSENPVDLVWTRCGHIAIRGQ	60
QY	70	LIRAYDYRHFSECFIPREAGEAYDFIPTSDPPEQRFALANOVGMPPVDLEN	129
DB	61	LIRAYDYRHFSECFIPREAGEAYDFIPTSDPPEQRFALANOVGMPPVDLEN	120
QY	130	RIDELACSLIESLRPOGQCNTEDEYAPFIRIFMLAGLPEEDIPHLKYLTDQMPDPG	189
DB	121	RIDELACSLIESLRPOGQCNTEDEYAPFIRIFMLAGLPEEDIPHLKYLTDQMPDPG	180
QY	190	SMTPFAEKALDYLLITIEORRQKPGCTDAISIVANQVGRPTSDPAKRMCGLLVGG	249
DB	181	SMTPFAEKALDYLLITIEORRQKPGCTDAISIVANQVGRPTSDPAKRMCGLLVGG	240

Oy	250	LDYVNFNFSEFMEFLAKPEHROELTEPPEFLIPACCELLRRSEVLVNDGHIILSDVEFHG	309
Db	241	LDYVNFNFSEFMEFLAKPEHROELTEPPEFLIPACCELLRRSEVLVNDGHIILSDVEFHG	300
Oy	310	VOJLKGDDIILPOMLSGIDEXKNAKCPMHVDFSRCKVSHTFHGSJLCLQHLARREIV	369
Db	301	VOJLKGDDIILPOMLSGIDEXKNAKCPMHVDFSRCKVSHTFHGSJLCLQHLARREIV	360
Oy	370	TLKELVTRIPDFSIAPGAOIOHKSGIVSGVALPLVNDPATTKAV	414
Db	361	TLKELVTRIPDFSIAPGAOIOHKSGIVSGVALPLVNDPATTKAV	405

RESULT 2

Y04126

ID Y04126 standard; Protein; 446 AA.

DC Y04126;

DT 11-JUN-1999 (first entry)

DE Bacterial and mammalian chimeric cytochrome P450 protein.

KW Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;  
oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;  
bioremediation; environmental pollutant.

OS Synthetic.

PN W09008812-AL.

PD 25-FEB-1999.

PF 17-AUG-1998; U16979.

PR 20-AUG-1997; US-056754.

PA (OYRP) UNIV ROCHESTER.

PI Jones JP, Shimoji M,  
WPI; 99-190131/16.

DR N-PSDB: X19916.

PT New P450 fusion proteins - comprising a portion of a bacterial  
cytochrome P450 protein and a portion of a mammalian cytochrome P450  
protein 24.

PS Claim 24; Page 6-8; 51pp; English.

CC The present sequence is a fusion proteins comprising a portion of a  
bacterial cytochrome P450 protein and also a portion of a mammalian  
cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or  
any compound having a carbon-hydrogen bond. The fusion protein can be  
used for hydroxylating a compound to be oxidised. It can also be used  
in the bioremediation of an environmental pollutant. Since the fusion  
protein is soluble, it can be subject to structural elucidation by X-ray  
crystallography for designing functional proteins. It can be readily  
expressed in soil bacteria to facilitate bioremediation.

Sequence 446 AA;

Query Match	54.1%	Score 1178:	DB 1:	Length 446:
Best Local Similarity	58.1%	Pred. No. 7.5e-114:		
Matches	255:	Conservative	30:	Mismatches 100: Indels 54: Gaps 10:

  

QY	10	NIAPLPVPHVPEHLVDFDMYNPSNLASGVQEAAMVLIQSNVPDLVWTRCNGCHVIATRCQ	69
Db	1	NIAPLPVPHVPEHLVDFDMYNPSNLASGVQEAAMVLIQSNVPDLVWTRCNGCHVIATRCQ	60
QY	70	LIREAYEDYRHSSSCFPFPRAGAYVQFIPISMPBPQROPRALANVYGAPVYDKLEN	129
Db	61	LIREAYEDYRHSSSCFPFPRAGAYVQFIPISMPBPQROPRALANVYGAPVYDKLEN	120
QY	130	RIQELACSLIESLRQGGCNFTEDYEAERPPIRIFMLIAGLPREDIPHLKYLLDQMTRPDQ	189
Db	121	RIQELACSLIESLRQGGCNFTEDYEAERPPIRIFMLIAGLPREDIPHLKYLLDQMTRPDQ	180
QY	190	SMTPFAEAEALYDYLIPITTEORQKRCPT-----DAISIVANGVQNGRP--ITTSDEAKKM	241
Db	181	SMTPFAEAEALYDYLIPITTEORQKRCPT-----DAISIVANGVQNGRP--ITTSDEAKKM	240
QY	242	CGLLVGLGDLVYVNFLEFSMEFLAKSPE--HROELIER-----DELIPA	283
Db	241	AVDLFGACTETTTSTTLRKALLLLLKHPETAYAVQGEIERIVIGRNNSPCMDGRSHAPYDA	300
QY	284	ACEELLRRFSLVADG--RIITSDYEFHGVQVLKKQDQILL-----PQMLSG-	326

Db 301 VVHEQVRIDTLPTSLPHAVYCDIKFNLYLLPKSTTLLISLSTVLDHNRKFRPEMDRH 360

QY 327 --IDERKACACMHNDPFSQKQKSHHTFGHSHLSCGLQRLAREITVYLKEWTLRPFDSA 384

Db 361 HFLDDEGN-----FKSKT-FPFSAGKRICVGEALDAGMELFLFTSLDQNFNLKSLV 412

QY 385 PQAQIQHKSGIVSGVALP 403

Db 413 DPKNLD-PTTPVNGFASVP 430

RESULT 3  
ID R60777  
AC R60777; standard; Protein; 396 AA.  
DT 21-JUN-1995 (first entry)  
DE Mycinamicin IV hydroxylating protein.  
KW Mycinamicin; hydroxylase; macrolide; antibiotic; Micromonospora;  
KW Micromonospora griseorubida.  
OS Micromonospora griseorubida AT11725CN3.  
PN J06253853-A.  
PD 13-SEP-1994.  
PE 09-MAR-1993; 047638.  
PR 09-MAR-1993; JP-047638.  
PA (ASAH ) ASAH KASEI KOGYO KK.  
WP1: 94-328997/41.  
DR N-PSDB: 073674.  
PT DNA encoding a protein having mycinamicin IV hydroxylating  
PT activity - for prodn. of mycinamicin, a macrolide antibiotic  
PS Claim 1; Page 12-14; 23pp; Japanese.  
CC The amino acid sequence of a protein having mycinamicin IV hydroxylase  
CC (M) activity. The gene encodes a protein of 396 a.a. The DNA was  
CC obtained from the macrolide antibiotic-producing bacterium Micromonospora  
CC griseorubida AT11725CN3/pts507. The gene was isolated from the plasmid  
CC ppts507. The protein encoded by this plasmid can be used to produce  
CC mycinamicin IV in pts507-deficient Micromonospora strains.  
SQ Sequence 396 AA;

Query Match	17.0%	Score 370;	DB 1;	Length 386;
Best Local Similarity	28.8%;	Pred. No. 4.6e-30;		
Matches 106;	Conservative 57;	Mismatches 167;	Indels 38;	Gaps 8;
OY	WIATGQILREAYEYEDYRHFSSECPRI-----PRAGEAYDFIPTSMDEPQROFPA	113		
Db	43 WLVT-----YELVRAVLGDGRVGRGSHMTREDPRRPEVVKGLSMDEPHSRLLR	95		
OY	114 LANQVGNRPVYDKLENRIQIOLACSLIESLPQG--CNFTEDYAEFPPIRIFMLAGLPEE	172		
Db	96 LVVKAFPTARRRESLPPRRAREIHMLVDMAATGCPADLVAMFARQLPVAVICELLGVPSA	155		
OY	173 DIPHLKYLTDQMTGRDGS-----TFARAKALDYLLPIIEORRKGSTALISV	223		
Db	156 D-----HDFRTKSGAFNSTAEVTAEBMODPAADAYAMODLLDRRKEETDDIVSAL	208		
OY	224 ANGVGNRPITSDERKRCGGLLVGLDPTVNFLSFSMEFLAKSEPHROELIERELLPA	283		
Db	209 VOARDQOSLSQELLDLAIGLVAGVESTPTQIDPFYLLMTBRELRRQLLDRELIPS	268		
OY	284 ACEELLRRFSL---VADGRILTSDEYFEGVQLKKDQILLPMLSGLDERKNAACSMHNF	340		
Db	269 AVEELTRNVPRLGVGTAPRPAVEDTLKGVITRAEPRVLAATGAANRDOAPRDAADRIDV	328		
OY	341 SROKVSHTTFGSHGISGONLARREILVTTLKEMWTRIPDFISA--PGAOIQHKSG--IVSG	398		
Db	329 DRTQNHNGFGHGVNHCISGARPLARVELQVALLEVLLQRLGRIGLGRIPETQLRMSBEMLLRG	388		
OY	399 VOALPPLVM 406			
Db	389 PLLEPLVM 396			

## RESULT 4

R77867  
ID R77867 standard; protein; 408 AA.  
AC R77867;  
DT 13-NOV-1995 (first entry)  
DE S. clavuligerus ORF10 product.  
KW clavulinate acid; clavulinate; antibiotic; beta-lactamase-inhibitor.  
OS Streptomyces clavuligerus.  
PN CA2108113-A.  
PD 09-APR-1995.  
PF 08-OCT-1993; 108113.  
PR 08-OCT-1993; CA-108113.  
PI (UVAL-) UNIV ALBERTA.  
AU Aldoo KA, Jensen SE, Paradkar AS;  
DR WPI: 95-207301/28.  
NR N-PSDB: 091580.  
PT Clavulinate acid biosynthesis enzymes and corresp. DNA - useful for PT biosynthesis of the antibiotic in Streptomyces hosts which do not naturally produce clavulinate  
PS Claim 32; Fig.19; 41pp; English.  
CC A 15 kb fragment S. clavuligerus NRRL 5741 genomic DNA (091580), extending downstream from pcbc, included 10 ORFs encoding the CC enzymes required for clavulinate biosynthesis. The ORF10 CC product (R77867) showed high similarity to cytochrome P450-type CC enzymes from other Streptomyces spp.  
SQ Sequence 408 AA;

Query Match 15.58; Score 338; DB 1; Length 408;  
Best Local Similarity 29.38; Pred. No. 1e-26;  
Matches 123; Conservative 59; Mismatches 184; Indels 54; Gaps 18;

16 PHVEHLEFDEDMYNPSLSAGVOEAMAVLQESNVPDLWTRNGGH-WIATRGQLIREA 74  
14 PAYPMHRKCPVD---PPQLAGLSQRAASRYT-----LM--DGSQVWLVSAGARAV 62  
75 YEDYRHS-SECEPIP-----REAGEAYDEIPTSMDEPEQROFRA-----LANOV 118  
63 LGDRRFRTAVTSAFCPMLRTSOLVRANPESASF--RMDPDQHSRLASMLTRDFLARRA 120  
119 VGM-PVYDKLENRIQELACSLIESLRPOGQCNFTEDYAEPPIRIFMLAGLPREDIPHL 177  
121 EALRPAYREL---LDEILGLVKGERR---VDLVAGLTIPVPSRYITLLFCAGDDRREFI 174  
178 K-----YLTDDMTRPDGSMTFAKEALVDYLPIIEQRORRKGTDALSTIVANGVNRPI 233  
175 EDRSAVLIDKRYTPE--QVAKARDELGYRLREVEERIEENPGTDLISRLVIDOVRGHL 231  
234 TSDEAKRMCGILLVGLDTVNVNPLSFSMEFLAKSPENHROELIERPELIPACEELLRRFS 293  
232 RVEEIVPNCRLILVAGHGTTSQASLSLSLTLPDELALGRLEDPALLPKAVEELLRRFS 291  
294 LVADG--RIITSDEYFHGVOLKKGDDIILPOMLSGLDERKNACPMHVDTSRQKVSHTTEG 351  
292 IVONGLAFAAIVEDQDLDDVLIRAGEGVVLSAGNRDETIVLPDPKRVDDADARHILAFG 351  
352 HGSMLCGOHLAR---REIITYLKEMLTRIPDSIA--GCAIQHKSGIVS-CYQALPLVW 406  
352 HGMHOCIGOMLARELELLELAVLRKM--PGARLAVFEELDFRHEVSYGLALPVTW 408

RESULT 5  
W33274  
ID W33274 standard; protein; 587 AA.  
AC W33274;  
DT 17-FEB-1998 (first entry)  
DE S. fradiae tylosin biosynthesis gene product tyIH.  
KW Tylosin; biosynthetic gene product; production; antibiotic; tyIH.  
OS Streptomyces fradiae.  
PN US672497-A.  
PD 30-SEP-1997.  
PF 21-DEC-1995; 575843.  
PR 12-MAY-1989; US-351350.  
PR 21-MAR-1986; US-842330.

PR 25-JUL-1986; US-890670.  
PR 24-FEB-1987; US-018237.  
PR 06-AUG-1991; US-742222.  
PR 28-JUL-1993; US-107232.  
PR 17-FEB-1994; US-198672.  
PR 21-DEC-1995; US-575843.  
PA (ELIL ) LILLY & CO ELI.  
PI Cox KL, Fishman SE, Hersherberger CL, Seno ET;  
DR WPI: 97-488660/45.  
NR N-PSDB: T58686.  
PT DNA encoding Streptomyces fradiae tylosin biosynthesis gene products PT - for increasing tylosin production in Streptomyces spp.  
PS Claim 20; Columns 21-38; 38pp; English.  
CC The present sequence is the Streptomyces fradiae tylosin CC biosynthetic gene product tyIH, useful to increase the production of the antibiotic tylosin in Streptomyces spp.  
SQ Sequence 587 AA;

Query Match 15.48; Score 336.5; DB 1; Length 587;  
Best Local Similarity 29.68; Pred. No. 2.6e-26;  
Matches 106; Conservative 62; Mismatches 169; Indels 21; Gaps 11;

63 WIATRGQLIREAYEDYR---HFSSECEPIPREAGEA-YDFIPTSMDEPEQRFALANOV 118  
143 WLISRODHVRLALADPRYSIH-PAKLRLSPDCEAEASRSLLTLDPPDHALGHFTPE 201  
119 VGMPEVYDKLENRIQELACSLIESLRPOG-QCNFTEDYAEPPIRIFMLAGLPREDIPHL 177  
202 FGLRRVRRELRSVQIYVGLDLDLTARDDEADLADFLPMATQVICHLDIPYEDRYF 261  
178 KYLTDDMTRPDGSMTFAKEALVDYLPIIEQRORRKGTDAL--STVANGVNRPIITS 235  
262 QERTEQATRPAGEBALLELDLRYDLRLISGTRGSGGMLGSMVA--QARGGGISH 319  
236 DEAKRMCGILLVGLDTVNVNPLSFSMEFLAKSPENHROELIERPELIPACEELLRRFS 295  
320 ADVLDNANVLLAAGHETPASTVMTSVLVLLQHPAMRRLTYVNPGLLPVAVDELLARYLS-I 378  
296 ADG--RIITSDEYFHGVOLKKGDDIILPOMLSGLDERKNACPMHVDTSRQKVSHTTEG 353  
379 ADGLRRSATADIIEIDGHTIRAGDCLVFLAANRDEAVFSEPAFDIHRARRHVAFCYG 438  
354 SHLCLGOMLARELITVLEKMLTRIPDSIAP-----GCAIQHKSGIVSQVQALPLVW 406  
439 PHOCLGOMLARELEVALGAVLERLP--ALRPITDVAGELRLKSDSA-VFGYVELPVAN 493

RESULT 6  
R11349  
ID R11349 standard; protein; 406 AA.  
AC R11349;  
DT 05-JUN-1991 (first entry)  
DE Cytochrome enzyme P450SUI.  
KW Cytochrome P450; P450SUI; P450S02; herbicide resistance.  
OS Streptomyces griseolus.  
PN WO9105561-A.  
PD 21-MAR-1991.  
PF 27-AUG-1990; 004785.  
PR 11-SEP-1989; US-405605.  
PR 12-JAN-1990; US-464499.  
PR 23-AUG-1990; US-569781.  
PA (DOPO ) DU PONT DE NEMOURS CO.  
PI Dean G, Harder PA, Lelo KJ, Lichtner FT, Odell JR;  
PI O'Keefe DP, Omer CA, Romesser JA;  
DR WPI: 91-102077/14.  
NR N-PSDB: 011126.  
PT DNA encoding cytochrome P450 enzymes - and electron donating PT iron sulphur proteins, used to confer herbicide resistance to PT plants and microorganisms  
PS Claim 13; page 151; 224pp; English.  
CC This cytochrome P450 enzyme, P450SUI is expressed alongside the iron sulphur protein Fes-B, by a DNA sequence contained in a recombinant

CC plasmid. Host *Streptomyces* species are transformed with the plasmid  
CC and are used to coat a plant seed to transform the plant. The res-  
CC ultant transformants are resistant to herbicides.

Sequence 406 AA;

Query Match	15.4%	Score 336	DB 1	Length 400
Best Local Similarity	26.6%	Pred. No.	1.6e-26	
Matches 105	Conservative	74	Mismatches	192
			Indels	24
			Gaps	11

```

OY      31 PSNLSAGVQ--EAMAVLOESNPDLWTRCNGCH-WIATRCOLIREAVEDYR----- 79
      ||| | | : : | : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      17 PSNNSCPYQLPDGYAQLRDTPGCLHRTVLYDGRQAMVVTTHKAARKLLIGDPRLLSSNRTDD 76

```

```
Oy      80 HESSECPYIP--REAGEAYDIPTSDPPEQROFALANOVGMVVDKLENIQELACS 137  
          :|: | |: |: |||| | : : : : :|:  
Db      77 NEPATSPREAVRESPQAF----IGLDPPHEGTRRRMTTSETFYVKRIKGMPVEEVELVHG 132
```

Oy 138 LHSLRPGC-QCNFTEDYAEPPRIRIMLAGLPEEDIPHLKYLTDOMTRDGSMTFAEA 196  
::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 133 FLEMLAGPTADLVSGFALVPSPMTCRLGLPYADHFFQDASKRLVSTDAQSALTA 192

Qy 197 KEALYDYLPIIEQRROKPGTDAI-STVANGVYNGRPITSDEAKRMCGLLVGLDPTVN 255  
: | | : : : : | | : | : | :  
Db 193 RNDLAGYDLGLITQOTEPGAGLVGALVADOLANGE-IDREELLSTAMLLITAGHETTAS 251

OY 256 FLSFSMEFLAKSPHEHQELIERPELIPACCEELIRFSL--VADGRILTSYEFGHGYLK 313  
| | | | | : | | | | : : | | | | : :  
Db 252 MTSLSVITLLDHPQYAALRADRSVPGAVEELLRYLATADIAGRVAATADIEVEGHILR 311

QY 314 KGDILLPMLSGLDERKNACPMHVDFSRQKVSHTTFGGSHLCLGQHLARREIIVTLKE 373  
| : : : : | : | | | | | | | : | |  
Db 312 AGEVIVVANSIARRGTVEEDPDALDIHRSARRHLLAFGGVHQCLGQNLRLLEVLINA 371

```

QY      374 WLTRIPDFSIA-PGAQIQHSG-IVSGVQALPELV 406
      : : : : : : : : : : : : : : :
Db      372 LMDRVPTLRLAVPVEQLVLRPGTTIQGVNELPEYV 406

```

RESULT	7
R11350	
ID	R11350 standard; Protein; 403 AA

DT	05-JUN-1991 (first entry)
DE	Cytochrome enzyme P450SU2.
KW	Cytochrome P450; P450SU1; P450SU2; herbicide resistance

PN	W09103561-A.
PD	21-MAR-1991.
PF	27-AUG-1990;

PR 12-JAN-1990; US-464499.  
PR 23-AUG-1990; US-569781.  
PA (DUPO ) DU PONT DE NEMC

PI O'Keefe DP, Omer CA, Romesser JA;  
DR WPI; 91-102077/14.  
DR N-PSDB; Q11127.

PT iron-sulphur pro  
PT plants and micro  
PS Claim 15; page

CC sulphur protein FeS-A, by a DNA sequence contained in a recombinant  
CC plasmid. Host *Streptomyces* species are transformed with the plasmid  
CC and are used to coat a plant seed to transform the plant. The res-

CC ultant transformant  
CC See also Q1126.  
SQ Sequence 403 AA;

Query Match 15.18; Score 328; DB 1; Length 403;

Best Local Similarity 28.7%; Pred. No. 1,1e-25;  
Matches 104; Conservative 64; Mismatches 165; Indels 30; Gaps 13;

```

QY 63 WATRGOLIREAVEYDRHSESEC----PFIPREAGEAYDPIPT--SMDBPEQORERALA 115
    | : | | : | | : | | : | | : | :
Db 52 WLVTNRHQDRAVLGSPR-FSDAHRTGFPFLTAGGREILITNPFLRMDDPREHARLRML 110

```

QY 116 NYVGMPPVDKLENNIOELACSLIESRP-QGOCNFTEDEYAEPPIRIFIMLAGLBEEDI 174  
: : : : : : : : : : : : : : : :  
Db 111 TADFIVKKKEAMREPVQRLLDDLVDRNTTGRISADLYEFALPLPSLVICLLIGVYEDH 170

```
Oy 175 PHL-----KLTLDQMRPDGSMTEFAEKKEALYDLPIITEQRKRPGTDAIS-IVANGQVN 229
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 171 AFQERSRVLLTLPSTPE---EVRADDELLEYLRLARTKRERPDAAISRLVARGELD 227
```

QY 230 GRPITSDEAKRMCGILLVGGIDTVAVNPLSFSMEFLAKSPENROELIERPELLIPACEELL 289  
| : | | | : : : | : | : | | |  
Db 228 DPGIAT----MGRLLLVAGGELDDQTQIATMGLLRNPDQLARLRAEPALVKGAVEELL 282

QY 290 RRSLSVADG--RILTSDEFFHGVLKGGDOLLPMLSGLDERKNACP--MHVDFSRÖKV 345  
|::|||::|::|:  
Db 283 RYLITIVHNHVPRIATEDVLIGRTIAGEGVLC--MISSANRDAAEVEFPGGDDLDVARDAR 340

Dy 346 SHITFGHSHLCLQHLARRRIIVTLAKEMWTRIPDFSTA-  
PGAOIQHKSGI-VSGYQALP 403

Dd 341 RHVAFGFSVNQLCQRPLARVELDIALETLLRRRLDLRAVP  
NHEEPIFRGDMAITYGNSLR 400

QY	404	Lvw	406
		:	↓
Db	401	Iaw	403

RESULT	8
W54389	
ID	W54389 standard; Protein; 411 AA

DT 18-AUG-1998 (first entry)  
DE *Acinomadura hibisca* polyketide synthase protein 10.  
KW Multienzyme; infection; fungi; yeast; gram-positive bacteria; virus

OS Actinomadura hibisca.  
PN WO9811230-A1.  
PD 19-MAR-1998.

PR 13-SEP-1996; WO-U14791.  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO  
PI Dairl T, Oki T;  
00 007081470

DR	N-PSDB; V266
DR	P-PSDB; W543
PT	Actinomadura
DT	radiata

PS	Disclosure
CC	The Actino
CC	a multienz
CC	aibudshar

CC is an antibiotic useful against systemic fungal infections caused by *Candida albicans*, *Aspergillus fumigatus* and *Cryptococcus neoformans*. It is also active against a wide variety of fungi and yeasts, some

Sequence 411 AA;

Query Match	15.08; Score 327; DB 1; Length 411.
-------------	-------------------------------------

Matches	114;	Conservative	61;	Mismatches	199;	Indels	58;	Gaps	16;
4	TIQSNNALAPRPHVDEHLVDEDMYNP----	SNLSAGVQEAVALQIESNVPDLVWTRCN	59						

60 GGH-WIATRGQLIREAYEDYRHFSSEC---PIPREAGEAYDFTPTS-----MDPPE 107

Db 51 GDHAWTVTRVDREFTVSD--RRFSKAVTRPGADP-----LIPMORGSKSLVIMDPPE 102  
QY 108 QROFALANOVGMPPVVDKLENRIQELACSLIESLRPOG--OCNFTEDYAEPPPIRIFMLL 166  
Db 103 HTBMKIVYSRAFTARVGMRAHVRDLTSGVDEMEVHGEPADLIALHLALPLPTVICEM 162  
QY 167 AGLPEDIPHLKYLTQDQTRPDGSMTPAEAKE-----ALYDYLPIITEQRQKGTDAI 220  
Db 163 LGVPEDRRPQDWTDRML--TIGAPALAOADEIKAAVGRGLYLAELLDAKTAAPADLL 221  
QY 221 SIYANGOVNRPITSDEKRRMCGLLVGGDLTVVNFLSFSEFLAKSEHGOELLEREL 280  
Db 222 SLTSRAHAD--GLSEELLTFGMLTLAGYHTTAATHTSVYHLRPSRYARLRDEPSG 280  
QY 281 IPACEELLRRFSVLADG----RLTSPDEPHGVOLKKGDOILLPOMLSGLDERKNACPM 336  
Db 281 IPAABEELL--RYGQIGGACAGIRIAYEDVEVGGTLVRGEAVIPLFNANANDPEVADPE 339  
QY 337 HVDESROKVSHTTEGSHLCLGQHLARREIIVTLKEMLTIRIPDSIA--PGAQIOHKSQI 395  
Db 340 ELDGRTDNPHIALGHGHIYCLGAPLARLELOVLETLVERTPALRLAIDADITWRPGL 399  
QY 396 V-SGVQALPLTW 406  
Db 400 AFARPDALPIAW 411  
RESULT 9  
ID W11585 standard; Protein; 398 AA.  
AC W11585;  
DT 02-APR-1997 (first entry)  
DE Streptomyces pristinaespiralis snbf gene product.  
KW Streptogramin B: antibiotic; biosynthesis; pristinamycin;  
KW virginiamycin; dipecolic acid; cyclodeamination; papa; snbf; snbf;  
OS Streptomyces pristinaespiralis.  
PN MO9601901-A1.  
PD 25-JAN-1996.  
PF 04-JUL-1995; F00889.  
PR 08-JUL-1994; FR-008478.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Barriere JC, Blanc V, Crouzet J;  
PI Debussche L, Paris JM, Thibaut D, Bamas-Jacques N;  
PI Dutruc-Rosset G, Famechon A;  
DR WPI: 96-097631/10.  
DR N-PSDB: T58555.  
PT New streptogramin B deriva. useful as antibiotics - produced by new  
PT mutants of Streptomyces having altered genes for streptogramin B  
PT biosynthesis  
PS Example 1: Page 113-114; 146pp; French.  
CC The papa gene of S.pristinaespiralis is involved in the biosynthesis  
CC of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for  
CC pristinamycin 1A. Upstream of the papa gene, on the complementary  
CC strand, is the snbf gene coding for 3-hydroxypicolinic acid-AMP ligase.  
CC The region between these two genes was sequenced and two open reading  
CC frames were identified. The first (pipa) decodes to an amino acid  
CC sequence with homology to ornithine cyclodeaminase from Agrobacterium  
CC tumefaciens. The pipa gene product is likely to catalyse the cyclo-  
CC demethylation of lysine, leading to production of dipecolic acid.  
CC Mutations in the pipa gene were shown to affect dipecolic acid  
CC synthesis but not the synthesis of 3-hydroxypicolinic acid. The second  
CC open reading frame (snbf) could be decoded to give a product with  
CC homology to hydroxylases of the cytochrome P450 type.  
CC Disruption of the pipa and snbf genes can be used to produce  
CC strains of S.pristinaespiralis which are unable to produce the  
CC antibiotic pristinamycin I but which may be able to produce new,  
CC modified forms of it.  
SQ Sequence 398 AA;

Query Match 14.8%; Score 323.5; DB 1; Length 398;  
Best Local Similarity 29.0%; Pred. No. 3.1e-25;

Matches 106; Conservative 59; Mismatches 159; Indels 41; Gaps 12;  
QY 74 AVEEYRH-----FSSECPFIPREAGEAYDPIPTSDPEORQOFALANOVGM 121  
Db 36 AFHYFRHADVLTVASDPCGVYSQSLRRLRGSOALSSEQLSYIDPPMRTLRLVSOAFTP 95  
QY 122 PVDKLENRIQELACSLIESLRPOG--NFTEDYAEPPPIRIFMLLAGLPEDIPHLKYL 180  
Db 96 RTVADLEPRVTELQGLLDVAV--DQDFDLVADFAVPLPVIVIAELLGVPRADRTLPRSW 153  
QY 181 TDQMT-----PGSMTPAEAKEALYDYLPIITEQRQKGTDAISIVA 224  
Db 154 SDRMLQMVADPADMQFDDADEDYQRLVKPEPMRMAHYIHDHTDRARPANDLSALV 213  
QY 225 NGOVNRPITSDEKRRMCGLLVGGDLTVVNFLSFSEFLAKSEHGOELLERPE--LIP 282  
Db 214 AARBERLTDEQIYVERGALLMAGHVSTSMILGTVLCLDHP--RKAARADRSLIP 271  
QY 283 AACCELLR-RPSLVADGRLTSDYEFHGVOLKKGDOILLPOMLS-GLDERKNACPMHYDF 340  
Db 272 ALIEVELRLRPITVMAVTTKQTVLAGTTIPAG--RMVPSLLSANHDEQVFTPRDIDL 330  
QY 341 SRQVSHTEFGHSHLCLGQHLARREIIVTLKEMLTIRIPDSIAPGAQIO--HKSQIVSGV 399  
Db 331 AREG-ROIAFGHGHYICGLAPLARLEGRILAEALFDRPPDSPDGAKLRYHROGLF-GV 388  
QY 400 QALPL 404  
Db 389 KNLPL 393  
RESULT 10  
ID R51368 standard; Protein; 410 AA.  
AC R51368;  
DT 24-NOV-1994 (first entry)  
DE Protein containing Cytochrome P450 SCA-2 activity.  
KW Cytochrome P450 SCA-2; Streptomyces carbophilus;  
KW treatment of hyperlipidaemia; drug preparation.  
OS Streptomyces carbophilus.  
FH Key Location/Qualifiers  
FT protein 1..410  
FT /label= cytochrome P-450 containing activity  
FT J06070780-A.  
PD 15-MAR-1994.  
PF 28-AUG-1992; 229969.  
PR 28-AUG-1992; JP-229969.  
PA (SANY ) SANKYO CO LTD.  
DR WPI: 94-128679/16.  
DR N-PSDB: Q61452.  
PT Cytochrome P450 SCA-2 gene - from Streptomyces carbophilus,  
PT useful in treatment of hyperlipidaemia  
PS Claim 1; Page 12-14; 18pp; Japanese.  
CC R51368 shows a protein having cytochrome P-450 activity. P-450  
CC SCA-2 can be prepared commercially for use in the preparation of a  
CC drug for the treatment of hyperlipidaemia.  
SQ Sequence 410 AA;

Query Match 14.7%; Score 319.5; DB 1; Length 410;  
Best Local Similarity 25.9%; Pred. No. 8.5e-25;  
Matches 93; Conservative 63; Mismatches 184; Indels 19; Gaps 7;  
QY 63 WIATRGQILREAYEDYR-----HSSSECPFIP--REAGEAYDFIPTSDPEORQOF 112  
Db 56 WVVTKHEARLRLADPRSSDRLHADFPATSPRKAFFQSGSPAF----IGMDPEHGTRR 111  
QY 113 ALANOVGMPPVVDKLENRIQELACSLIESLRPOG--OCNFTEDYAEPPPIRIFMLLAGPE 171  
Db 112 RMTISEFTVKRIKGRPDVERIVHGFIDDMLAAGPTADLVSOFAIPVYSWVICHMLGVPY 171  
QY 172 EDIHLKYLTDQMTRPDGSMTFAEAKKALYDYLPIITEQRQKGTDAISIVANGOVNGR 231

```
Dh 172 ADHEFDODASRLVQAVDADASAAARDFFERYLDGLITFKLESEPGTGLGLVTHQADG 231
QY 232 PITSDEAKRMCGLLVGGIDTVVNFSLFSMEFLAKSPENROELIERPELIPACEELLRR 291
Dh 232 EIDRAELISTALLLVGHSHETASMTSLVTTLEHPDQNALRADPSLVGAVPELLRV 291
QY 292 FSL--VADGRILTSYERHGVQAKGDOILLPOMLSGIDERKNACPMHWDESRQKVSHTT 349
Dh 292 LAIDINGGRATADIEDIGOLIRAGEGVIWNSIANDSSVFENPRLDVHRSARHILS 351
QY 350 FGHSHLICGHLARREIYVTKEMLTRIPDFSTA-PGAQIQHKS-G-IVSGVQALPLVW 406
Dh 352 FGCVHOCIGONLARLEVLIVLPDRIPRLAVPEVQLLTRPGTTIGVNEPLVWTM 410

RESULT 11
ID R38309 standard; Protein; 412 AA.
AC R38309;
DT 04-DEC-1993 (first entry)
DE Sequence of the P450-soy protein.
KW P450soy; soy; gene; oxidation; haem protein.
OS Streptomyces griseus ATCC 13273.
PN M09312235-A.
PD 24-JUN-1993.
PF 16-DEC-1992; U10885.
PR 16-DEC-1991; US-807001.
PI (DUPO ) DU PONT DE NEMOURS & CO E. I.
PI Omer CA, Sariaslani FS, Trower MK;
DR WPI: 93-214178/26.
DR N-PSDB; Q45569.
PT Constitutive expression of P450 SOY (SOYC) and ferredoxin soy
PT (soy) in Streptomyces - used for oxidn. of organic chemicals
PS Example; Figure 2; 45pp; English.
CC Cytochrome P450soy was purified from S. griseus ATCC 13273. Two
CC similar forms of P450soy were isolated. P450soy-delta, is derived
CC from P450soy by in vitro proteolysis during isolation. One of the
CC tryptic peptide fragments of cytochrome P450soy and of of the
CC P450soy-delta protein were subjected to automated degradation. The
CC N12 terminal sequences are given in R38306 and R38307. A mixture of
CC oligos that consist of possible DNA sequences that could encode the
CC AAS RGYNHCL of the tryptic peptide was made. It consists of the
CC sequences in Q43290-23. The oligo mixture was end-labeled and used
CC to probe the EMBL4 library of S. griseus DNA. Hybridizing plaques
CC were isolated and a 4.8kb SacI DNA fragment was isolated from one
CC clone that hybridised to the oligo probe mixture. As segment of the
CC 4.8kb fragment was sequenced and found to contain an ORF. Within
CC this ORF was a section that corresp. exactly to the AA sequence
CC determined from the cytochrome P450soy tryptic peptide (see Q45569,
CC R38309). The gene encoding the P450soy protein was called soyc. Five
CC nucleotides downstream from the stop codon for soyc another ORF was
CC identified. This ORF encodes an apparent ferredoxin-like protein.
CC The gene was designated soyb and the protein ferredoxin-soy.
SQ Sequence 412 AA;
```

```
Query Match 14.5%; Score 315; DB 1; Length 412;
Best Local Similarity 27.2%; Pred. No. 2,5e-24;
Matches 84; Conservative 56; Mismatches 163; Indels 6; Gaps 5;
```

```
QY 103 MDPEQOFALANQVGMVVDKLENRIQELACSLIESLRPGG-QCNFTEDVAPPEPTIR 161
Dh 105 VDDPEHNTORMMLIPTFSVKRIGALRPRIQETVDRLDAMERGCPAPALVSALPVPBM 164
QY 162 IFMLAGLPEEDIRPHLYLDQMTRRPGSGMTFAKAEILDYLLPIEQRQKGTDAIS 221
Dh 165 VICALLGVPAADNAFFERSQORLGRPADVDVNRABELEUYGALIDRKRAEFGDGLD 224
QY 222 IVANGVNGRPITSDEAKRMCGLLVGGIDTVVNFSLFSMEFLAKSPENROELIERPEL 281
Dh 225 ELIHRDPRDGPVREQLVAVRIVILLINGHETTAMISLGFTTLLSHPEQLAALRAGSTST 284
QY 282 PAACEELLRRPSLVADG--RLTSDEYFHVQVQAKGDOILLPOMLSGIDERKNACPMHVD 339
```

```
Dh 285 AVVEEELL-RFLSLAEGIQRLATFEDMEVDGATIRKGEVVFSTSLINRDADVPERAETLD 343
QY 340 FSRQKVSHTTEFGHSHLICGHLARREIYVTKEMLTRIPDFSTA-PGAQIQHKS-G-IVS 397
Dh 344 WDRPARHHLAFGCVHOCIGONLARLEVLIVLPDRIPRLAVPEVQLLTRPGTTIGVNEPL 403
QY 398 GVCALPLVW 406
Dh 404 GLIDLPLVAM 412
```

```
RESULT 12
ID R14724 standard; Protein; 404 AA.
AC R14724;
DT 28-JAN-1992 (first entry)
DE 6-hydroxylase encoded by eryf gene.
KW C-6 hydroxylation; erythromycin; 6-deoxyerythromycin; antibiotics;
KW saccharopolyspora; cytochrome P450 monooxygenase; ss.
OS Saccharopolyspora erythraea.
PN M09116334-A.
PD 31-OCT-1991.
PF 16-APR-1991; U02600.
PR 18-APR-1990; US-510483.
PI (ABBO ) ABBOTT LABORATORIES.
PI Weber JM;
DR WPI: 91-339744/46.
DR N-PSDB; Q14548.
PT New 6-deoxyerythromycin derivs. - are antibiotics with increased
PT acid stability, produced by cultivation of saccharopolyspora.
PS Disclosure; Fig 3; 56pp; English.
CC The eryf gene encodes the 6-hydroxylase component of the cytochrome
CC P450 monooxygenase system responsible for the hydroxylation of 6-de-
CC oxyerythronolide B to erythronolide B. Interruption of this step
CC results in the formation of deoxyerythromycin A and new derivatives
CC useful as antibiotic which have better stability against acids that
CC the corresponding erythromycins. Interruption of the reaction can
CC be effected by an insertion into the eryf gene of a plasmid, gene
CC replacement or chemical or light-induced mutagenesis.
CC See also Q14549.
SQ Sequence 404 AA;
```

```
Query Match 13.6%; Score 296.5; DB 1; Length 404;
Best Local Similarity 24.2%; Pred. No. 2e-22;
Matches 90; Conservative 74; Mismatches 173; Indels 35; Gaps 10;
```

```
QY 63 WIATRGQILREAYEDVNHFS-----BCPLIPREAG--EAYDFIPTSM--DPEEQ 108
Dh 40 WLVTGYDEAKAALSDLRSLSPDKKKYGVVEF-PAVLGPEEDVRNYFATNMGTSDDPTH 98
QY 109 RQFALANQVGMVVDKLENRIQELACSLIESLRPGG-QCNFTEDVAPPEPTIRIFMLAG 168
Dh 99 TRLRKLVSOEFTYKRVKRAMRPVEQIYAEILDEGDSGVVDIYVRANPLRPIVICELLG 158
QY 169 LPEDIDPHLYLD-----QMTRPDGSMTFAEAKAEALDYLLPIEQRQKGTDAI 220
Dh 159 VDE-----KYRGEFGWSEELVMPERABQGAAREVNVNITLDVERKRRREPDDLL 212
QY 221 STVANGQ--VNGRPITSDEAKRMCGLLVGGIDTVVNFSLFSMEFLAKSPENROELIERP 278
Dh 213 SALIFVQDDDDGR-LSADELSTALVLLLAGFEASVSLIGITGYLLTHPDQALAVARDP 271
QY 279 ELIPACEELLRRPSL-VADGRILTSYERHGVQAKGDOILLPOMLSGIDERKNACPMH 337
Dh 272 SALPNAVEELIRYIALQETTTFRPAEDLEIGVAIPQYSTVLVANGAANDPQFPDPHR 331
QY 338 VDESKGVSHTTGHSNLCIGHLARREIYVTKEMLTRIPDFSTA-PGAQIQHKS-G-IVHKS 395
Dh 332 SDVTKTRGHLSTGQGHIFPCMGRLAKLEGEVALRALFGFRPALSLGTDADVYVRRSYL 391
QY 396 VSGVQALPLVWD 407
```

33 N-terminal modified daunomycin C-14 hydroxylase.





---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:38:05 ; Search time 66.92 Seconds  
(Without alignments)  
103.693 Million cell updates/sec

Title: US-09-246-451-12  
Perfect score: 2179  
Sequence: 1 TTTTQSNANLAPLPPHVE.....IVSGVALPLVMDPATTKAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 1676186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgnl\_7/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgnl\_7/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgnl\_7/ptodata/1/1aa/6.COMB.pep:\*  
4: /cgnl\_7/ptodata/1/1aa/PCTUS.COMB.pep:\*  
5: /cgnl\_7/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	371.5	17.0	416	US-09-320-878-18	Sequence 18, Appl
2	346	15.9	403	5212296-9	Patent No. 5212296
3	336	15.4	406	5212296-6	Patent No. 5212296
4	315	14.5	412	US-08-102-863-11	Sequence 11, Appl
5	315	14.5	412	PCT-US92-10885-11	Sequence 11, Appl
6	277.5	12.7	419	US-09-335-409-8	Sequence 8, Appl
7	160	7.3	422	US-09-086-982-5	Sequence 5, Appl
8	160	7.3	422	US-08-653-650A-5	Sequence 5, Appl
9	160	7.3	422	US-09-096-982-8	Sequence 8, Appl
10	160	7.3	474	US-08-653-650A-8	Sequence 8, Appl
11	157	7.2	443	US-09-096-982-9	Sequence 9, Appl
12	157	7.2	443	US-08-653-650A-9	Sequence 9, Appl
13	155	7.1	422	US-08-396-218-2	Sequence 2, Appl
14	155	7.1	422	US-08-760-116-2	Sequence 2, Appl
15	135	6.2	512	US-08-194-981E-5	Sequence 5, Appl
16	129.5	5.9	382	US-09-320-878-7	Sequence 7, Appl
17	119	5.5	501	US-08-906-791-2	Sequence 2, Appl
18	117	5.4	513	US-08-948-664-6	Sequence 6, Appl
19	111	5.1	492	US-08-724-466B-2	Sequence 2, Appl
20	109.5	5.0	504	US-08-457-274A-25	Sequence 25, Appl
21	109.5	5.0	504	PCT-US95-05758-25	Sequence 25, Appl
22	108.5	5.0	490	US-08-201-118-7	Sequence 7, Appl
23	108.5	5.0	490	US-08-238-821B-7	Sequence 7, Appl
24	108.5	5.0	490	PCT-US95-05744-7	Sequence 7, Appl
25	106.5	4.9	490	US-08-201-118-3	Sequence 3, Appl
26	106.5	4.9	490	US-08-201-118-9	Sequence 9, Appl
27	106.5	4.9	490	US-08-238-821B-3	Sequence 3, Appl
28	106.5	4.9	490	US-08-238-821B-9	Sequence 9, Appl

29	106.5	4.9	490	PCT-US95-05744-3	Sequence 3, Appl
30	106.5	4.9	490	PCT-US95-05744-9	Sequence 9, Appl
31	105.5	4.8	496	US-08-313-075A-50	Sequence 50, Appl
32	105	4.8	497	US-08-724-466B-4	Sequence 4, Appl
33	98.5	4.5	472	US-08-622-166A-2	Sequence 2, Appl
34	98.5	4.5	472	US-08-622-166A-4	Sequence 4, Appl
35	98.5	4.5	498	US-08-457-274A-24	Sequence 24, Appl
36	98.5	4.5	498	PCT-US95-05758-24	Sequence 24, Appl
37	97.5	4.5	1026	US-07-998-003A-95	Sequence 95, Appl
38	97.5	4.5	1026	US-08-453-274B-95	Sequence 95, Appl
39	97.5	4.5	1026	US-08-453-695A-95	Sequence 95, Appl
40	97.5	4.5	1026	US-08-268-161A-95	Sequence 95, Appl
41	97.5	4.5	1026	US-08-453-702A-95	Sequence 95, Appl
42	97.5	4.5	1026	PCT-US93-1258B-95	Sequence 95, Appl
43	97.5	4.5	1026	PCT-US93-08071-95	Sequence 95, Appl
44	97.5	4.5	1203	US-07-998-003A-103	Sequence 103, App
45	97.5	4.5	1203	US-08-453-274B-103	Sequence 103, App

ALIGNMENTS

```
RESULT 1
US-09-320-878-18
: Sequence 18, Application US/09320878A
: Patent No. 6117659
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanie C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
: APPLICANT: TANG, Li
: TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
: FILE REFERENCE: 300622002120
: CURRENT APPLICATION NUMBER: US/09/320,878A
: EARLIER FILING DATE: 1999-05-27
: EARLIER APPLICATION NUMBER: CIP OF 09/141,908
: EARLIER FILING DATE: 1998-08-28
: EARLIER APPLICATION NUMBER: CIP OF 09/073,538
: EARLIER FILING DATE: 1998-05-06
: EARLIER APPLICATION NUMBER: CIP OF 08/846,247
: EARLIER FILING DATE: 1997-04-30
: EARLIER APPLICATION NUMBER: 60/119,139
: EARLIER FILING DATE: 1999-02-08
: EARLIER APPLICATION NUMBER: 60/100,880
: EARLIER FILING DATE: 1998-09-22
: EARLIER APPLICATION NUMBER: 60/087,080
: EARLIER FILING DATE: 1998-05-28
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 18
: LENGTH: 416
: TYPE: PRT
: ORGANISM: Streptomyces venezuelae
US-09-320-878-18
```

Query Match 17.0%; Score 371.5; DB 3; Length 416;  
Best Local Similarity 28.6%; Pred. No. 1.4e-30;  
Matches 98; Conservative 62; Mismatches 162; Indels 21; Gaps 7;

```
QY 76 EDYHFSSECFIRREGEADVDFPTSMDDPEQOFRALANOVYGMPPVVDLENRIQELA 135
DB 72 KDWK--NSTPLTEAEALNLMLES--DPRHRTLRVLAREFTMRVVELLRVEIV 127
QY 136 CSLLEST--RPOGOCNFEDYAEPPFIRIFMLAGLPEEDIPHLKYLTDOMTRPDGSMTF 193
DB 128 DGLVDAMLAARDGRADLMESLAWPLPITVISELGVPEPDAAFRVWDAPVFPDDPAQA 187
QY 194 AEAKFALYILPITEORRQPGTDAIS--IVANGOVNCRPITSDEAKRMCGLLVGLDT 252
DB 188 QTAAEMKSGYLSRLDSKRCODGEBLLSAIVRTSDEDSRLTSELLGMHAILLVAGHET 247
```

```

OY      253 VNFNFSFSEFLAKSPENHOELIERPELLPAACEELLRRFSFIVAQRILTSDFE----- 307
          ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 308
DB      248 TVNLANGVYALSHPDQALALRADMTLLDGAVEKL-----YESPVESATYRRPEVEY 302

OY      308 --HGVLKKGDQIILPOMLSGLDERKNACPMVDSRSKOVSTTTGGHSHLGLGHLAR 365
          ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 366
DB      303 DLDGVIVIPAGDVLVLADAHRTPRFPDPHRRFDLRDPAIGHAGHGTHFCIGAPLARL 362
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 367
OY      366 EIVTLKEWLRTRIPDFS--IAPGAQIOHSGIVSGVQALPLW 406
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 407
DB      363 EARIARVALLERCPDLALDVPGLVWYVNPATRIKALPIW 405

RESULT 2
5212296-9
PATENT NO. 5212296
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSEY, JAMES A.
TEPPERMAN, JAMES M.
TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
CYTOCHROMES
NUMBER OF SEQUENCES: 19
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/569,781
FILING DATE: 23-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 464,499
FILING DATE: 12-JAN-1990
APPLICATION NUMBER: 405,605
FILING DATE: 11-SEP-1989
SEQ ID NO.9:
LENGTH: 403
5212296-9

Query Match      15.9%; Score 346; DB 5; Length 403;
Best Local Similarity 29.2%; Pred. No. 6,1e-28;
Matches 106; Conservative 63; Mismatches 164; Indels 30; Gaps 13.

OY      63 WATRGQLIREAYEDYRHESSSEC-----PFIREAGEAYDFTPT--SMDPEQOROFALA 115
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 116
DB      52 WLWTRHODRAVLGDPR--FSADAHRTGFPFLTAGGHEIGTNPFLRMDDPEHARLRML 110
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 117
OY      116 NQVQMPVYDKLENTIOELACSLIESLR--QOCQNTEDYEAARFPRITMLAGCPREED 174
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 175
DB      111 TADFTVKKKEARPEQRLADLVPMTTGRTSADLVTEFALPLSLVLCILLGVYEDH 170
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 171
OY      175 PHL-----KYLTQOMTRPDGSMFAEAKELAYDYLPIIQRQRKPGTDALS--IVANGVN 229
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 230
DB      171 AFGQERSRVLLLRSPPE--EVRAAQDBELRLYLARLARTKRERPDALLISLTVARGELD 227
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 228
OY      230 GRPITSDEAKRKCGLLVGGDLTVVNFLSFSNEFLAKSPENHOELIERPELLPAACEELL 289
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 290
DB      228 DTQIAT-----NGRLLLVAGHETTANMTALSTVLVLRNPDQALRLARAEPAIVKGAVEELL 282
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 291
OY      290 RRFSLVADG--RILSDYEFHGVOLKKGQDQIILPOMLSGLDERKNACP--MHVDSRSKOV 345
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 346
DB      283 RLTVIVHNSVPRIATEDVILIGRTTIAAGEGVLC--MISSANDAEVFPPEGDDLDVARDAR 340
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 347
OY      346 SHTTGGHSHLGLGHLARREITVTLKEMLRTRIPDFSIA--PGAQIOHSGI--VSGVQALP 403
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 404
DB      341 RHVAFGFGVHCGTGLQPLRAVELQIAIETLLRLRLRLRLAVPHEEIPFRGDMAIVGVHSIP 400
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 405
OY      404 LVM 406
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 407
DB      401 IAW 403
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 408

```

```

TEPPERMAN, JAMES M.
: TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
CYTOCHROMES
: NUMBER OF SEQUENCES: 19
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/569,781
FILING DATE: 23-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 464,499
FILING DATE: 12-JAN-1990
APPLICATION NUMBER: 405,605
FILING DATE: 11-SEP-1989
SEQ ID NO.:6
LENGTH: 406
5212296-6

Query Match 15.4%; Score 336; DB 5; Length 406;
Best Local Similarity 26.6%; Pred. No. 6,9e-27;
Matches 105; Conservative 74; Mismatches 192; Indels 24; Gaps 11.

QY 31 PSNLSAGVQ--EAMAVLQGSNVVDLWTRCNGH-WIATRGQLIREAVEYDR----- 79
Db 17 PSNRCPYQLPDGYAQLDTPGLRHRTVLTYDGRQAAMVVTKHEARKLGADPRLSSNRDD 76
QY 80 HFSHCPCIP--REAGEAYDFIPTSMDBPEEQOPALANOVMGVVVKLERIOELACS 137
Db 77 NFPAISPFEEAVRESPOAF---IGLDPEHGTRRRMTISEPTVARIKGMREVEEVYHG 132
QY 138 LIESLRPGQ-QCNLFEDYAEPFPITIFMLLAGLPREDILHLKYLLTDOKTRPGSMTFAEA 196
Db 133 FLDEMLAGAPTAADVSQLPVPMSVICRLLGVPADHEFFQDASKRLVAQSDAQSALTRA 192
QY 197 KEALDYALPIETIORRQRGTDAI-SIVANGOVNRPITSDEAKRMCGLLVGGLDTVVN 255
Db 193 RNDLAGLYDLGITQQTGEBAAGVALVAOJLANE-IDREBLISTAMILLIAGHETTAS 251
QY 256 FLFSMEFLANSPEERQELIERBELIPAACEEELRRFSL--VADGRILTSDYEFGVOLK 313
Db 252 MTSLSVITLLHPEDYVALRADRSLVPGAVEELLRYLAIDADIAGRVAADAIEVBGHILR 311
QY 314 KGDQILLQOMLSGLDERKNACPMHYDFSROKVSHTFGGHSILCIQONLARREITYTKE 373
Db 312 AGEQIVVANSTIANRCGYIEDPDALDIRSAHHHLAPFGVHOCIGONLARLELEYILNA 371
QY 374 WLTRIPIDFSIA-PGAOIQHKSQ-IVSGVOALPLVW 406
Db 372 LMDRVPLTLANVPVEQLVLRPCTTIQGVNELPVTW 406

RESULT 4
US-08-102-863-11
: Sequence 11, Application US/08102863
: Patent No. 5466590
GENERAL INFORMATION:
APPLICANT: SARTIASANTI, STMA
TITLE OF INVENTION: CONSTITUTIVE
TITLE OF INVENTION: EXPRESSION OF P450SOY
TITLE OF INVENTION: AND FERRODOXIN-SOY IN
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS
ADDRESSEE: AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS

```



SEQ ID NO 8  
LENGTH: 419  
TYPE: PRF  
ORGANISM: Sorangium cellulosum  
US-09-335-409-8

Query Match 12.7%; Score 277.5; DB 3; Length 419;  
Best Local Similarity 24.0%; Pred. No. 9, 8e-21;  
Matches 101; Conservative 67; Mismatches 178; Indels 75; Gaps 14;

QY 6 QSNANAPLPYVPEHLYVPEDMYNSNLSAGVQ-----AMAVLOSNNVDLYWTRCNG 61  
DB 3 QEQANQSEKRP-----AFDFKPPAP-----GYAEDFPFAIERLRRA-TPIFYWD--EGR 48  
QY 62 HWIATRGQ-----LIREAYEDYHFSSECFIRREAGEAVDFIPTSMDPEQR 109  
DB 49 SWVLTTRYHDVSAVFRDERFVAVSREEMSSAEYSATP-----ELSDMKKYGIFGLPREDHA 104  
QY 110 QFRALANQVGMFVVDKLENRIQELACLISLIRPOQCNETDYAEPPRIRIFMLLAGL 169  
DB 105 RVKRLVNPSTSRALDILRLAERIQRTVDQLLDARSGQEEFVDVADYAGIPRAISALLKV 164  
QY 170 PEEDIDHLYKLTQMTTRPQDSMTFAEKKALYLLP-----I 207  
DB 165 PAF-----CDEKRRRFSAT-----ARALGVLPVQVDETKTLVASVTEGLALHDV 212  
QY 208 IEORROKP-GTDAISIVANGVGRPTTSDCAKRMGCLLVGLDTPVNFLEFSMEFLAK 266  
DB 213 LDERRRNPLENDVLTMLQAEADGSRSTELVALYGAITADTTIYLAFAVNLRL 272  
QY 267 SPEHROELIERPELLPACCELLRFSLVADG--RILTSDYEFHGVOLKKGOI--LLPQ 322  
DB 273 SPEALELVKAEPLMRNALDEVLRFDNILLRIGTVREFARODLEYCGASIKKGEMVFLIPS 332  
QY 323 MSLGDERKNACPMHVDFSQKVSHTTFGSHLCLGCHLAREIIVTLKEMLTRIPDES 362  
DB 333 ALR--GCTVFSRDPVDVDRDGTASLACGCPVPCGVSILARLEAIVAGTIRPREPMK 390  
QY 383 I 383  
DB 391 L 391

RESULT 7  
US-09-096-982-5  
Sequence 5, Application US/09096982  
Patent No. 5962293  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEE, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/096,982  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Goltick, Mary E.  
REGISTRATION NUMBER: 34829

REFERENCE/DOCKET NUMBER: 22727/00131  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 216-622-8458  
TELEFAX: 216-241-0816  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-096-982-5

Query Match 7.3%; Score 160; DB 2; Length 422;  
Best Local Similarity 23.6%; Pred. No. 2, 1e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 QFAMAVLOSNNVDLYWTRCNGHWIATRGQOLIREAYEDYRHFSSE--CPFIPTREAGEAY 96  
DB 31 REAGPVV-EVNAPE-----AGPRAWITDDALAREVLADPRFVKGPDLPATAMRGVDDGL 83  
QY 97 D-----FIPTSMRPREQORFALANQVGMFVVDKLENRIQELACSLIESTL-----R 143  
DB 84 DIPVPELRPEPTLIAVDGEDHRRRLRIHAPFNPRRLAERTDRIATAIDLTELADSSOR 143  
QY 144 PQQCNETDYAEPPRIRIFMLLAGL-----EDIDPHLYKLT-----TDQMTTRPD 188  
DB 144 SGPRALIGGFANHFPLLYICELLSGVPTDPMARAVGLNALGSGQSGAGDGTDA 203  
QY 189 GSMTEAKAEALYLLPIIEORRQKPGTDAISIVANGVGRPTTSDCAKRMCGILLVG 248  
DB 204 GDVPTSALEST--LLEAVHAARRKDTMTTRVLYERAQAEFGVSDDDLVYMITGLIPA 261  
QY 249 GIDTVYNPLFSMEFLAKSPEHROELIERPELLPACCELLR-----PSYVADGRILTS 303  
DB 262 GHDITGSFLGF--LAEVLARLADADGDAISRVEALRHNRPVPSL--WFEPAAT 315  
QY 304 DYEFHGVOLKKGOIILLPMLSGDERKNACPMHVDFSQKVSHTTFGSHLCLGCHLA 363  
DB 316 EYVIRGVRLPRGAPVLDLEGTTDGRHNHDAPRIAPDRSRRLRTFGSGPHNICIGEOLA 375  
QY 364 RREIIVTLKEMLTRIPDESIA--PGAQIQ--HNSGIVGVQALPLVW 406  
DB 376 QLESFTMIGVLSRFPQARLAVYEELRWCKRKGAGOTARLTDLR-VW 420

RESULT 8  
US-08-653-650A-5  
Sequence 5, Application US/08653650A  
Patent No. 5976830  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEE, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/653,650A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:



```

? APPLICATION NUMBER: US/08/653,650A
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Goldick, Mary E.
? REGISTRATION NUMBER: 34829
? REFERENCE/DOCKET NUMBER: 22727/00131
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 216-622-8458
? TELEFAX: 216-241-0816
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 474 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-653-650A-8

Query Match 7.38; Score 160; DB 2; Length 474;
Best Local Similarity 23.6%; Pred. No. 2.5e-08;
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 OEAMAVLQESNVDPDLVWTRCGNGHMITRGQLIREAYEDYRHSS--CPPIRAGEAY 96
DB 83 REAGPVV-EVNAP-----AGGRAVITTDALAREVLADRFVKGPDLPATMAGVDDGL 135
QY 97 D-----FPTSMDEPREGORFALANQVGMFVVDKLENRIQELACSLIESL-----R 143
DB 136 DIPBELRPFTLLAVDQEDHRLRIHAPAFNPRLAEPTDRIAADRLTELAOSSDR 195
QY 144 POGCQNTEDYAEPRIRIFMLAGLP-----EEDIPLKYL-----TDQMPRPD 188
DB 196 SGEPAELIGFAVHFPLVLCGLGVPTDPMAREAVGVKALGLGPGSAGGDTGDP 255
QY 189 GSMFAFAKALDYLPITIDROROKPGTDAISVANGVNGRPTDSEAKRMCGLLVG 248
DB 256 GVDPTSALESLL--LLEAVNAARKKDTRTMTRVLYERAQAEFGVSDDOLVYMTGLIFA 313
QY 249 GLDVTNVFLFSMEFLAKSPENROELIERPELLPAACELLLR-----FSLVADGRILTS 303
DB 314 GHDTGSFLGF---LLAEVLAGRLADADGDALSRFVEALRNHPRYTSL--WRFAAT 367
QY 304 DYEHGVOLKKGOILLPOMLSGLDERKNACPMHNDPSROKVSHTTFGHSILCLOHLA 363
DB 368 EYVLRGVRLPRGAPVLVDIEGTNTDGRHNDAPAFHPRDRSRRLTFGDGRPHYCIGQLA 427
QY 364 RREIYVLKEMLRIRIPFSLA-PGAQIQ--HKSGIVSGVQALPLVW 406
DB 428 QLESRTMIGVLRSPQARLAVPYEELRMCRKGAQTARLTDLR-VW 472

RESULT 11
US-09-096-982-9
? Sequence 9, Application US/09096982
? Patent No. 5962293
? GENERAL INFORMATION:
? APPLICANT: Strohl, William R.
? APPLICANT: Dickens, Michael L.
? APPLICANT: Desanti, Charles L.
? TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
? NUMBER OF SEQUENCES: 9
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: CALFEE, HALTER & GRISWOLD
? STREET: 800 Superior Avenue, Suite 1400
? CITY: Cleveland
? STATE: Ohio
? COUNTRY: USA
? ZIP: 44114-2688
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/096,982
? FILING DATE:
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Goldick, Mary E.
? REGISTRATION NUMBER: 34829
? REFERENCE/DOCKET NUMBER: 22727/00131
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 216-622-8458
? TELEFAX: 216-241-0816
? INFORMATION FOR SEQ ID NO: 9:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 443 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? US-09-096-982-9

Query Match 7.28; Score 157; DB 2; Length 443;
Best Local Similarity 23.6%; Pred. No. 4.6e-08;
Matches 95; Conservative 50; Mismatches 210; Indels 48; Gaps 13;

QY 39 OEAMAVLQESNVDPDLVWTRCGNGHMITRGQLIREAYEDYRHSS--CPPIRAGEAY 96
DB 52 REAGPVV-EVNAP-----AGGRAVITTDALAREVLADRFVKGPDLPATMAGVDDGL 104
QY 97 D-----FPTSMDEPREGORFALANQVGMFVVDKLENRIQELACSLIESL-----R 143
DB 105 DIPBELRPFTLLAVDQEDHRLRIHAPAFNPRLAEPTDRIAADRLTELAOSSDR 164
QY 144 POGCQNTEDYAEPRIRIFMLAGLP-----EEDIPLKYL-----TDQMPRPD 188
DB 165 SGEPAELIGFAVHFPLVLCGLGVPTDPMAREAVGVKALGLGPGSAGGDTGDP 224
QY 189 GSMFAFAKALDYLPITIDROROKPGTDAISVANGVNGRPTDSEAKRMCGLLVG 248
DB 225 GVDPTSALESLL--LLEAVNAARKKDTRTMTRVLYERAQAEFGVSDDOLVYMTGLIFA 282
QY 249 GLDVTNVFLFSMEFLAKSPENROELIERPELLPAACELLLRFSLV--ADGRILTSDE 306
DB 283 GHDTGSFLGF---LLAEVLAGRLADADGDALSRFVEALRNHPRYTSLMRFATEVY 339
QY 307 FHGVOLKKGOILLPOMLSGLDERKNACPMHNDPSROKVSHTTFGHSILCLOHARE 366
DB 340 IRGVRLPRGAPVLVDIEGTNTDGRHNDAPAFHPRDRSRRLTFGDGRPHYCIGQLAOLE 399
QY 367 IIVTLKEMLRIRIPFSLA-PGAQIQ--HKSGIVSGVQALPLVW 406
DB 400 SRITMIGVLRSPQARLAVPYEELRMCRKGAQTARLTDLR-VW 441

RESULT 12
US-08-653-650A-9
? Sequence 9, Application US/08653650A
? Patent No. 5976830
? GENERAL INFORMATION:
? APPLICANT: Strohl, William R.
? APPLICANT: Dickens, Michael L.
? APPLICANT: Desanti, Charles L.
? TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
? NUMBER OF SEQUENCES: 9
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: CALFEE, HALTER & GRISWOLD
? STREET: 800 Superior Avenue, Suite 1400
? CITY: Cleveland
? STATE: Ohio
? COUNTRY: USA
? ZIP: 44114-2688
? COMPUTER READABLE FORM:
```



```

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: PatentIn Release #1.0, Version #1.30
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/08/653,650A
7 FILING DATE:
8 CLASSIFICATION: 435
9 ATTORNEY/AGENT INFORMATION:
10 NAME: GOLPICK, Mary E.
11 REGISTRATION NUMBER: 34829
12 REFERENCE/DOCKET NUMBER: 22727/00131
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: 216-622-8458
15 TELEFAX: 216-241-0816
16 INFORMATION FOR SEQ ID NO: 9:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 443 amino acids
19 TYPE: amino acid
20 STRANDEDNESS: single
21 TOPOLOGY: linear
22 MOLECULE TYPE: peptide
23
24 US-08-653-650A-9

```

Query Match	7.2%	Score 157;	DB 2;	Length 443;
Best Local Similarity	23.6%	Pred. No. 4.6e-08;		
Matches 95;	Conservative 50;	Mismatches 210;	Indels 48;	Gaps 13;

QY 39 QEAAVAILOESVPLDVTMTCNGGHWITRCQGLIEAEVDYRHFSS--CFFITPEAGEAY 96  
 Db 52 REAGVY-EVNAF-----AGPAMVITTDALAEVLADREYKVDPLDPA7PAMRGVDDGL 104  
 QY 97 D-----FIPSTMDPPEOROFALANOVVGMFVDKLENRIOLCSTLESF-----R 143  
 Db 105 DIPVELRPFLLIADVGDGHDHRRRLRIRHAPAFNPRRLAEPTDRIALADRLITELDSSDR 164  
 QY 144 POGGCPNFEDYAEFPPIRIFMLLGLP-----EEDIPHLKYL-----TDQMPRPD 188  
 Db 165 SGEPAELLIGGAHYHFPPLVLTICELGVPVTPDPMAREAVGVKLALGLGPGSAGSGDGTDP 224  
 QY 189 GSMFPAEAKELLYUFLPIIROROKPCTDAISIVANGOVNCRITSDPAKRMCGILLYG 248  
 Db 225 GDVDPDTSALESF--LLEAVHAAKKDKTRITMTVLYERAQAEFGSVSDQLVMTGTLIPA 262  
 QY 249 GLDTPVNFLESTMEFLAKSPENHOELIERPELLIPACEELLRRFSLY--ADGRILTSDE 306  
 Db 283 GHDTTGSFLGF---LLEAVLGRLLAADADGALISRFVEALRHHNRPPTYLLMRAAREVY 339  
 QY 307 FHGYOLKKGDDIILPQMLSGDERKKMACPMHVDFSROKVSHTTTFGHSHLCLGQHLARE 366  
 Db 340 IRGVRLPFGAVPLVDIEGTNDGRRHNDAPHAHFHPDRPSRRRLTFGDGPYHCIGLOLAOLE 399  
 QY 367 IIVVLKEMLRIPDESTA-PCAQI--HKSSVSGVALPLVW 406  
 Db 400 SRTMIGVLRSEFPQARLAVPYLELRMCRKGAGQTRLDLP--VW 441

RESULT 13  
 US-08-396-218-2  
 : Sequence 2, Application US/08396218  
 : Patent No. 5695966  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: INVENTI, Augusto  
 : APPLICANT: BREME, Umberto  
 :  
 : APPLICANT: COLOMBO, Anna L  
 : APPLICANT: HUTCHINSON, Charles R  
 :  
 : APPLICANT: OTTEN, Sharee  
 : APPLICANT: SCOTTI, Claudio  
 :  
 : TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN  
 :  
 : NUMBER OF SEQUENCES: 4  
 :  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESS: NIKAIIDO, MARNEI,STEIN, MURRAY & ORAM

```

1 STREET: 655 Fifteenth Street, N. W., Suite 330 - G
2 STREET: Street Lobby
3 CITY: Washington
4 STATE: DC
5 COUNTRY: USA
6 ZIP: 20005-5701
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: Floppy disk
9 COMPUTER: IBM PC compatible
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11 SOFTWARE: PatentIn Release #1.0, Version #1.30
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/396, 218
14 FILING DATE:
15 CLASSIFICATION: 536
16 ATTORNEY/AGENT INFORMATION:
17 NAME: KITTS, Monica C
18 REGISTRATION NUMBER: 36,105
19 REFERENCE/DOCKET NUMBER: P1615-5002
20 TELECOMMUNICATION INFORMATION:
21 TELEPHONE: 202/638-5000
22 TELEFAX: 202/638-4810
23 INFORMATION FOR SEQ ID NO: 2:
24 SEQUENCE CHARACTERISTICS:
25 LENGTH: 422 amino acids
26 TYPE: amino acid
27 TOPOLOGY: linear
28 MOLECULE TYPE: protein
29
30 US-08-396-218-2

```

Query Match	7.1%;	Score 155;	DB 1;	length 422;
Best Local Similarity	23.8%;	Pred. No. 6.9e-08;		
Matches 97; Conservative	51;	Mismatches 202;	Indels 58;	Gaps 15

0Y	39	OEAMVAVLOESVPLDVMTKCNQGNHIAIRGOLIEAEVDYHNFSSCEPFIIPREGEAYDF	98
0Y	31	REAGVAV- EVNAP-----ACGPAMVITDDMLAREVLADPREFYVD-----PDLPAPAMRG	78
0Y	99	IPITSM D--PPEOROFALA-----NOVGMPEYD-----KLENILOELACSLIESL-	142
Db	79	VDDGDIDIVPELRLPTLLIAYDGEAHRRLRIHIAFNPRLAEKTRDIAIAGSLITELA	138
0Y	143	----PRQOCNFTEDYAEPPFIPIFMLLAGLP-----EEDIDHLKYL-----TDQ	183
Db	139	DASGSGKPAELIGFAFHAFFPLVLCELGVPVDDPAMAREAVSVLKALGIGQSGGSD	198
0Y	184	MTRPSSMTFAEAKFALVDYLPIILPEORQKPGIDAISIVANGOVNCRPTISDEAKMCG	242
Db	139	GTDPRGCVPTLSALESL--LLEAVHSAKRNDTPMTKRYLVERAQAEEGVSDDOLVYMIT	256
0Y	244	LLVGLGDTVNVFLTFSMEFLAKSPHENQOELTERPELIPACEBELLRFSLV--ADGRIL	301
Db	257	GLIFGHNHTTGSFLG---LLEAVIAGLADADDEDVASRVEEALRYHPPVPTTLKRA	313
0Y	302	TSDYEFHCVOLKCDQILLPOMLSGLDERKNACSPAHVDFSQKVSHTTFCHGSHLCLGH	361
Db	314	ATEVFIGGVRLPRGAPVLYDIEGTFWTDGRHNDHAHAFHDPSPRRRLTFEGDGPHYCIGQ	373
0Y	362	LARRIILYTLKEMLRIPIPDFSIA- PGADIQ--HKSIGVSGQALPLVW	406
Db	374	LAQLESRTMIGVLRSRFEARLAVYDEYDILMNCRCRGAQOATRLTELPR-WW	420

RESULT 14  
 US-08-760-116-2  
 : Sequence 2, Application US/08760116  
 : Patent No 5786190  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: INVENTI, Augusto  
 : APPLICANT: BREME, Umberto  
 : APPLICANT: COLOMBO, Anna L  
 : APPLICANT: HOTCHINSON, Charles R

APPLICANT: OTTEN, Sharee  
APPLICANT: SCOTTI, Claudio  
TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIKAIKO, MARIELEIN, MURRAY & ORAM LLP  
STREET: 655 Fifteenth Street, N. W., Suite 330 - G  
City: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/760,116  
FILING DATE: 3-DEC-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/396,218  
FILING DATE: 27-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: KITTIS, Monica C  
REGISTRATION NUMBER: 36,105  
REFERENCE/DOCKET NUMBER: P1615-6007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/638-5000  
TELEFAX: 202/638-4810  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-760-116-2

Query Match 7.1%; Score 155; DB 1; Length 422;  
Best Local Similarity 23.8%; Pred. No. 6.9e-08;  
Matches 97; Conservative 51; Mismatches 202; Indels 58; Gaps 15;  
DB 39 OEAMAVLQESNVPDLVWTRCGMHIAIRGQILREAVEDYHFSSECPFIPREAGEAYDF 98  
DB 31 REAGPVV-EVNAP-----AGGPAMVITDOLAREVLADPRFVAD-----PDLPAMARG 78  
QY 99 IPTSMQD--PPROQFRALA-----NQVGMFVVD-----KLENRIQELACSLIESL- 142  
DB 79 VDDGIDIPVPELRFPTLIVADGEAHRRLRIHAPAFNPRRLAERTDIRIAIAGRLTLELA 138  
QY 143 ---RPGQCFNTEDEVAEPPPIRIFEMLLAGP-----EEDIPHLVYL-----TDQ 183  
DB 139 DASRSCKPAELIGFAYHNPFLVYICELGVPTDPMAREAVYLAKLGCGPQSGGD 198  
QY 184 MTRPDGSMTEFAKEALYDYLPIIEQROKPGTDAISIVAMGVNRPITSDAKRMCG 243  
DB 199 GTDPAGGVPPDSALESL--LLEAVHSARNDTPMTFVLYERAAGFSGVSDDDLVYMIT 256  
QY 244 LLLVGLDTPVNNFLSFSNEFLAKSPENROELIERPELIPACCELLRRFSLV--ADGRIL 301  
DB 257 GLTFAGHDTGSGFLGF---LLAEVLARLADADEDAVSREVEBALHYHPVPYTLWRA 313  
QY 302 TSDVEFGVQLKKGDQILLPQMLSGLDERKACPMHVDFSROKYSHTTFGSHLCGQH 361  
DB 314 ATEVTTIGVRLPRGAPVAVLDGNTGCRHNDADAHAFHDPDSRRRLTFGGDPHYCIGEQ 373  
QY 362 LARREIIVTLKEMLTIRIPDESIA-PGAQIQ--HKSGIVSGVQALPIVW 406  
DB 374 LAQLESMTMIGVLSRFPPEARLAVPYDELKRCRGAQTARLTLP-VW 420

RESULT 15  
US-08-194-981E-5  
Sequence 5, Application us/08194981E  
Patent No. 5886157  
GENERAL INFORMATION:  
APPLICANT: GUENGERICH, F. Peter  
APPLICANT: GUO, Zuyu  
APPLICANT: SANDHU, Punam  
APPLICANT: GILLAM, Elizabeth M. J.  
TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF  
TITLE OF INVENTION: HUMAN  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
STREET: Suite 1200, 127 Peachtree Street, NE  
City: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303-1811  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/194,981E  
FILING DATE: February 10, 1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Elizabeth Selby  
REGISTRATION NUMBER: 38,298  
REFERENCE/DOCKET NUMBER: 22000.0022  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 688-0770  
TELEFAX: (404) 688-9880  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 512 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
FRAGMENT TYPE: N-terminal  
US-08-194-981E-5

Query Match 6.2%; Score 135; DB 2; Length 512;  
Best Local Similarity 22.4%; Pred. No. 1.2e-05;  
Matches 95; Conservative 67; Mismatches 158; Indels 104; Gaps 21;  
DB 27 DMVPSNLSGVQAMAVLQESNVPDLVWTRCGMHIAIRGQILREAVEDYHNF----- 81  
DB 107 DLYFTLLISNGQSMFSF-----PD-----SGPVAAAR-----RLAONGLSFSIASD 149  
QY 82 ---SSCEPFIPEAGEAYDFIPTSMQD---PPROQFR-----ALANOVGMFVVDKLENR 130  
DB 150 PASSTSCYLEHYSKEAEVLISTQELMAGRGHFNPRIVYVSTWVICAICGRRIIDN 209  
QY 131 IQELACSLIESLRPGQCFNTEDEVAEPPPIRIFEMLLAGLPEEDIPHLKYLTDOMTRPDGS 190  
DB 210 HQEL-LSLV-----NINNNNGE-----VGSNGNPAERFIPILRYL-----PNPS 246  
QY 191 M-TFAEAKELDYDYLPIIEQROK-----KQETDAISIVANGQ-----VNGRPITSD 237  
DB 247 LMAEKDLINEKFSYSPQMKVAKVKKTFEKGIRITTD--SLIEHQEKOLDENANVOLSD 304  
QY 238 AKRMCGLLVG-GIDTPVNNFLSFSMEFLAKSPENROELIE-----RPEL----- 280  
DB 305 KIINIYVLDLGAGFDYTTATISMSIMYLVANPRVQRKIQELDTVIGRSRRPRLSDSHL 364  
QY 281 -IPACEELLRRFSLV--ADGRILTSDEYFHVQLKKGDQILLPQMLSGLDERKACPM- 335

Db 365 PYMEAFLETFRHSSFPETIPHSTRTDTSLKGFYIPKGRCVFNQWQINHDOKLWNP 424  
QY 336 -----MHVDSROKV---SHTFGSHLCLGOHLARREIYTLKEMLTRIPDFSIAP 385  
Db 425 EELPERFLTPDGATDKVLSEKVIIFGMRKRCIGETVARWEVFLLAILLQRY-EFSVPL 483  
QY 386 GAQI 389  
Db 484 GVKV 487

Search completed: October 4, 2000, 12:38:07  
Job time: 21424 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: October 4, 2000, 12:36:41 ; Search time 73.09 Seconds  
(without alignments)  
350.531 Million cell updates/sec

Title: US-09-246-451-12

Perfect score: 2179

Sequence: 1 TTEIQSNANLALPPHPVE.....IVSGVALPLVMDPATKAV 414

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR-64:\*  
2: pir1:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2169	99.5	415	1	campor 5-monooxyg
2	378	17.3	410	2	cytochrome P450meg
3	374	17.2	393	1	cytochrome P450 Rv
4	372	17.1	396	1	cytochrome P450 yj
5	370	17.0	397	1	cytochrome P450 my
6	366	16.8	398	1	cytochrome P450 Rv
7	352.5	16.2	404	2	cytochrome P450 -
8	346	15.9	403	2	cytochrome P450 10
9	342.5	15.7	428	1	cytochrome P450 Rv
10	341.5	15.7	410	1	cytochrome P450 10
11	339.5	15.6	405	1	cytochrome P450 CV
12	339	15.6	410	1	cytochrome P450 cy
13	336	15.4	406	2	cytochrome P450 10
14	332	15.2	406	1	cytochrome P450 10
15	327.5	15.0	376	1	cytochrome P450 10
16	327	15.0	411	1	cytochrome P450 10
17	324	14.9	428	1	cytochrome P450 10
18	322.5	14.8	395	1	cytochrome P450 bl
19	319.5	14.7	410	2	cytochrome P450 sc
20	318.5	14.6	399	1	cytochrome P450 -
21	317.5	14.6	405	1	cytochrome P450 Rv
22	315	14.5	412	2	cytochrome P450 (s
23	311	14.3	310	2	cytochrome P450 (i
24	310	14.2	438	1	cytochrome P450 Rv
25	309	14.2	402	2	cytochrome P450 Rv
26	309	14.2	404	1	cytochrome P450 Rv
27	309	14.2	417	1	cytochrome P450 Rv
28	302.5	13.9	408	2	cytochrome P450 no
29	302.5	13.9	408	2	cytochrome P450 no

30	293.5	13.5	406	1	cytochrome P450 er
31	293	13.4	381	1	cytochrome P450 cy
32	291.5	13.4	433	1	cytochrome P450 Rv
33	290.5	13.3	337	2	cytochrome P450 hy
34	273.5	12.6	310	2	cytochrome P450 hy
35	271	12.4	414	1	cytochrome P450 Rv
36	268.5	12.3	411	2	cytochrome P450 Rv
37	267.5	12.3	386	2	cytochrome P450 Rv
38	259	11.9	401	1	cytochrome P450 Rv
39	259	11.9	412	1	cytochrome P450 Rv
40	255	11.7	411	1	cytochrome P450 Rv
41	254	11.7	406	3	cytochrome P450 Rv
42	240	11.0	396	1	cytochrome P450 Rv
43	233	10.7	433	2	cytochrome P450 Rv
44	231	10.6	398	1	cytochrome P450 Rv
45	229	10.5	402	1	cytochrome P450 Rv

## ALIGNMENTS

RESULT 1  
O4PSCP  
campor 5-monooxygenase (EC 1.14.15.1) cytochrome P450 101 - Pseudomonas putida plasm  
N:Alternate names: cytochrome P450-CAM  
C:Species: Pseudomonas putida  
C>Date: 30-Apr-1982 #sequence-revision 31-Dec-1993 #text-change 03-Mar-2000  
C:Accession: A25660; S34614; C60886; A00194  
J:Unger, B.P.; Gunsalus, I.C.; Sliagar, S.G.  
J: Biol. Chem. 261, 1158-1163, 1986  
A:Title: Nucleotide sequence of the Pseudomonas putida cytochrome P-450-cam gene and  
A:Reference number: A94678; MUID:86111751  
A:Accession: A25660  
A:Molecule type: DNA  
A:Residues: 1-415 <UNG>  
A:Cross-references: GB:M12546; MID:g151114; PIDN:AAA25760.1; PID:g151115  
R:Atamaki, H.; Koga, H.; Sagara, Y.; Hosoi, M.; Horinuchi, T.  
Biochim. Biophys. Acta 1174, 91-94, 1993  
A:Title: Complete nucleotide sequence of the 5'-exo-hydroxycampor dehydrogenase gene  
A:Reference number: S34613; MUID:93326643  
A:Accession: S34614  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-42 <ARA>  
A:Experimental source: PG31: ATCC 17453; CAM plasmid  
R:Romeo, C.; Moriwaki, N.; Yasunobu, K.T.; Gunsalus, I.C.; Koga, H.  
J. Protein Chem. 6, 253-261, 1987  
A:Title: Identification of the coding region for the putidaredoxin reductase gene fro  
A:Reference number: A60886  
A:Accession: C60886  
A:Molecule type: DNA  
A:Residues: 408-415 <ROM>  
R:Haniy, M.; Ames, L.G.; Yasunobu, K.T.; Shastri, B.A.; Gunsalus, I.C.  
J. Biol. Chem. 257, 12664-12671, 1982  
A:Title: Amino acid sequence of the Pseudomonas putida cytochrome P-450. II. Cyanogen  
A:Reference number: A00194; MUID:83030788  
A:Accession: A00194  
A:Molecule type: protein  
A:Residues: 2-55, 58-276, 'Q', 278-361, 'S', 363-407, 'N', 409-415 <HAN>  
C:Genetics:  
A:Gene: camC; CYP101  
C:Function:  
A:Description: catalyzes hydroxylation of campor to yield 5-exo-hydroxycampor; elec  
C:Superfamily: Pseudomonas plasmid campor 5-monooxygenase; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygen  
F:246-380/Domain: cytochrome P450 homology <CYP>  
F:358/Binding site: heme iron (Cys) (axial ligand) #status experimental

Query Match 99.5%; Score 2169; DB 1; Length 415;  
Best Local Similarity 99.5%; Pred. No. 2,9e+158;  
Matches 412; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY	1	TTETIOSNNANLAPLRPHVPEHLVPEDPMYNPSNLSAGVOEAAVLAQESVWPLVWTRCG	60
Db	2	TTETIOSNNANLAPLRPHVPEHLVPEDMNPNSLSAGVOEAAVLAQESVWPLVWTRCG	61
OY	61	GMIATRGOLIEAIVEDYHFSSECPFIIPREAGEAYDFIPTMDPEOROFALANOVYG	120
Db	62	GMIATRGOLIEAIVEDYHFSSECPFIIPREAGEAYDFIPTMDPEOROFALANOVYG	121
OY	121	MPVVDKLENRIOEACSLIESLRPOGOCNFTEDVAEPPRIEMULAGLPEEDIDHLKYL	180
Db	122	MPVVDKLENRIOEACSLIESLRPOGOCNFTEDVAEPPRIEMULAGLPEEDIDHLKYL	181
OY	181	TDOOMTPODSMTFAAKKELLYDYLPIITEQOROKGTAISIVANGOVNGRITSDCAKR	240
Db	182	TDOOMTPODSMTFAAKKELLYDYLPIITEQOROKGTAISIVANGOVNGRITSDCAKR	241
OY	241	MCGLLLVGGLDIPVWFLSPFSMEFLAKSPENHROELIERPILPAACEELLRRRSLVADGRI	300
Db	242	MCGLLLVGGLDIPVWFLSPFSMEFLAKSPENHROELLERPERIPAACEELLRRRSLVADGRI	301
OY	301	LTSDYEFHGVOLKKGDQILLPOMLSGLDERKNACPMHVDFSROKVSHTTFGHSHLCLGQ	360
Db	302	LTSDYEFHGVOLKKGDQILLPOMLSGLDERENACPMHVDFSROKVSHTTFGHSHLCLGQ	361
OY	361	HLAREEIIYTLKEWLTIRIPDFSIAFGAOIQHKSIGVSGVALPRLVWDATTATAY	414
Db	362	HLAREEIIYTLKEWLTIRIPDFSIAFGAOIQHKSIGVSGVALPRLVWDATTATAY	415

RESULT 2  
S39924  
cytochrome P450meg - *Bacillus megaterium* (ATCC 13366)  
C:Species: *Bacillus megaterium*  
A:Variety: ATCC 13368  
C:Date: 07-oct-1994 #sequence\_revision 23-Feb-1996 #text\_change 04-Mar-2000  
C:Accession: S39924; S32216  
R:Rauschenbach, R.; Isernhagen, M.; Noeske-Jungblut, C.; Boidol, W.; Stewert, G.  
Mol. Gen. Genet. 241, 170-176, 1993  
A:Title: Cloning sequencing and expression of the gene for cytochrome P450meg, the steroid  
A:Reference number: S39924; MUID:94049677  
A:Accession: S39924  
A:Molecule type: DNA  
A:Residues: 1-410 <RAU>  
A:Cross-references: EMBL:Z21972; NID:9288298; PIDN:CAA79965.1; PID:9288300  
A:Experimental source: ATCC 13368  
C:Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: Chromoprotein; heme; Iron; metalloprotein  
F:35/Binding site: heme iron (Cys) (axial ligand) #status predicted

```

Query Match      17.3%  Score 378;  DB 2;  Length 410;
Best Local Similarity 30.4%;  Pred. No. 2.1e-21;
Matches 110;  Conservative 53;  Mismatches 179;  Indels 20;  Gaps

OY 63 WIATGQILRAVEDYRHFSSE-----CPPIRAGEAVDFIP-TSDMDPEOROPALA 115
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 44 WNVFEYEDYKRVSLDYKHFSSYKRRTTITSVGTDSBESVSPEKIQITSDSDPHKRKSSL 1030
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

OY 116 NOVGMVYVDKLENNRIOLACSLIESLRPOGACFTFDVAEPPIRIFMLAGIPEDIP 175
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 104 AAATPRLSLQNMPERIQIADDELIGMDGTEIDIVASLSPLEIIVAMDLMGVPSDR 163
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

OY 176 HLKVTIDQMTFSDGSMTFAE-----AKELVYLIPIIQRQRKGTALSIIVANGOV 228
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 164 LEKKWVDLFLPFDRERKOEVDKLOVAAKXYOYLPIYVQKRLNPADDIISDLKSEV 223
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

OY 229 NGRPTSDSEAKRMCGLLLVGLDITVNVNLFESW-EFLAKSPHEOELIERPELIPACEE 287
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 224 DGEHPTDDEVRTMLILIGACVETTSHLNLSPLSLYDKOEYVQELHENDLTPQAVEE 263
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

OY 288 LLR-FESLIVADGRILTSDFEHHVQLKKDQDILLPOMLSGIDREKKNACPMHVDFSR-QKV 345
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```

Db	284	MARFENLIKIDRTYKEDNDLLGVELKEDDSVVMMSAANDEKEFDEPFTLNIHRPNK	343
QY	346	SHTTGGHGHSLCTGLHARRETIIVLKEMLTIPDPSTAPGAQIOHK--SGIYSGVAL	402
Db	344	KLTGTGNGPHFCLGAPLAEKIALTAFLKFKHIEAVPSFQLENTLDSATGOTLSTL	403
QY	403	PL 404	
Db	404	PL 405	

RESULT 3  
C70929  
cytochrome P450 RV1785c - Mycobacterium tuberculosis (strain H37Rv)  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Mycobacterium tuberculosis  
C:date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: C70929  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon-  
; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajadram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.  
A:title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: C70929  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-393 <COL>  
A:Cross-references: GB:A022021; GB:AL123456; NID:33250699; PIDN:CA17707.1; PID:e125  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV1785c  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase

Query Match	17.2%	Score 374	DB 1	Length 393
Best Local Similarity	28.7%	Pred. No. 4e-21		
Matches 110, Conservative	69	Mismatches 184	Indels 20	Gaps 10

[illegible]

RESULT 4  
B69851  
cytochrome P450 yj1B - *Bacillus subtilis*



Matches	108;	Conservative	69;	Mismatches	168;	Indels	34;	Gaps	8;
OY	55	WTRCN-----	GGHWIATRGQLIREATEDYRHSS-----	ECPCFIREAGEAYDIPT	101				
Db	26	WMRAQPVFRRNRNGIAAASYQAVIDAERQPELTSAGGIRPPDQPALP-----	PMI	76					
OY	102	SMDDPEQOFRALANQVGMPIVDKLENRIOELACSLIESLRPOGQCFETEDYAEPPIR	161						
Db	77	DMDDPAHLIRKKLVNAGTTRKKRVKDEASTIALCPTLIDACERECQCVRLAALPLDPA	136						
OY	162	IFMLLAGI-PEEDIPHLK-----	YLTDDMTRPDGSMTFAEAKALDYLIPIIQROR	213					
Db	137	VIGDMLGVRRPEQDQWELRMSDDLVTFLSHVSQEDFQITM-DAFAAYNDFRATIAARRA	195						
OY	214	KPGDTAISIVANGQVNGRPITSDCAKRMCGILLVGLDVTYVNFSLFSMEFLAKSEHNQE	273						
Db	196	DPDTDLVSLVSVSEVDGRSLSDDELMEFTLTLIGGDETTKHTLSTGTEQLLRNDDWL	255						
OY	274	LIERPELIPACCELTFRFSLVAD-GRITSPDYEHGQULKGGDITLIPQULSGDEKN	332						
Db	256	LORDPSLLPGALIEELMRKTAFAKMKRCRVLTADTFPHGATLAGEKMMLLFESANDEAVF	315						
OY	333	ACPMHWDSRQVSHHTTFGHSHLCLGQHLARRELIIVTLKEMLTIRIPDFSAPGAOI--Q	390						
Db	316	CEPEKFDVQQRPNNSHLACGFGHFFCLGQRLARLELSTLTERVLRRLPDLRLVADDSVPL	375						
OY	391	HKSGIVSGVALPLWMDPA	409						
Db	376	RPANFVSGLESMPVVFETPS	394						

[illegible]

```
DB      268 AKQASWMMANVEDVELSTLYKAGCAVAALQHSNTADPKYVDHPREIDFHKTSNPMSLGG 347
OY      352 HGSLLCLGOLHARREIIVTLKEWLTRIPDSIA-PGAOLIQHSG-IYSGOVALPLYW 406
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      348 HGAAHHMGCAQLVREMGTALGSLISRIPIALRPRAVEPRIKFLRGRLVPSLEALPLTW 404

RESULT      8
B35401
cytochrome P450 105B1 - Streptomyces griseolus
N:contains: oxidoreductase (EC 1.-.-.-)
C:Species: Streptomyces griseolus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 04-Mar-2000
C:Accession: B35401
R:Ommer, C.A.; Lenstra, R.; Little, P.J.; Dean, C.; Tepperman, J.M.; Leto, K.J.; Romess
J. Bacteriol. 172, 3335-3345, 1990
A>Title: Genes for two herbicide-inducible cytochromes P-450 from Streptomyces griseo
A:Reference number: A35401; MUID:90264332
A:Accession: B35401
A:Molecule type: DNA
A:Residues: 1-403 <OMD>
A:Cross-references: GB:M32239; NID:g153480; PIDN:AAA26825.1; PID:g153481; GB:M36481
C:Genetics:
A:Gene: CYP105B1
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenas
E:238-374/domain: cytochrome P450 homology <p45>
F:352/Binding site: heme iron (Cys) (axial ligand) #status predicted
```

```

Query Match 15.9% Score 346: DB 2: Length 403;
Best Local Similarity 229.2% Pred. No. 5.7e-19;
Matches 106; Conservative 63; Mismatches 164; Indels 30; Gaps 13;

OY 63 WATRGQLIREAYEDYRHFSSSEC-----PIIPREAGEAYDEIPT--SMDPEOROFALA 115
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 52 WLVTRHQVRAVLGDPR-FSADAHRTGPFLLAGREITIGNPTFLRMDDEHARLRML 110
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 116 NOVGMPPVADKLKNIIOGLACSLIESLRP-QGQCFTEDYAEPPFIRIFMLLAGIPEDI 174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 111 TADETVKKVEAMRPVQRLADLVDMRTTGTSALVTEFALPIPSLVICLLIGLGPEDH 170
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 175 PHL-----KYLTDMTRPDGSMTFEAKKALYDYLIPIEORRORGTDAIS-IVANGOVN 229
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 171 AFGQSRSVLLLTLRSTPE---EVRAAGDELELYLARLARTRRERDDALISRIVARGLD 227
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 230 GRPITSDAKRKRCGLLYLVGLDYVNVPLFSFMEFLAKSPERROELTEPELIPACEEL 289
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 228 DTQIAT-----MGRLLVAGHETTANMTALSTVLRLRPDOLARLRAPALVKGAVEL 282
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 290 RRFSLVAGG--RIILSDVEFHGVOLKKGDOLILPOMISGLBERKNAC--MHVPSRKY 345
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 283 RYLTIYHNGVRIATEDVLIGRTIAGEGVLC--MISSANRDAEVFGGDDLDVYARAR 340
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 346 SHTFEGHSHLCLGOLHARRELIYTLKEMLTRIPDFISA-BGADIQHKSGL-VSGVALP 403
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 341 RHVAGFGVHQCLOGPLARVELQIAIEFLRLRLPDLRLAVHEEIPFGDMATGVHSLP 400
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 404 LVW 406
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 401 IAW 403
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
F70729
cytochrome P450 RY2266 - Mycobacterium tuberculosis (strain H37RV)
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Mycobacterium tuberculosis
C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:accession: F70729

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

```



Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A:Accession number: A70500; MUID:98295987  
A:Reference: F70729  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-428 <COL>  
A:Cross-references: GB:277163; GB:AL123456; NID:93261610; PIDN:CAB00969.1; PID:e255175;  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV2266  
C:Superfamily: *Bacillus* cytochrome p450 CYP106; cytochrome p450 homology  
C:Keywords: oxidoreductase  
E:264-401/domain: cytochrome p450 homology <CYP>

Query Match	15.7%;	Score 342.5;	DB 1;	Length 428;
Best Local Similarity	26.5%;	Pred. NO. 1.1e-18;		
Matches 118;	Conservative 71;	Mismatches 188;	Indels 69;	Gaps 17;

[illegible]

```

RESULT      10
O4B56M
cytochrome P450 106 - Bacillus megaterium
N:Alternate names: cytochrome P450BM-1
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Bacillus megaterium
C:Date: 31-Mar-1992 #sequence
C:Accession: S07764; S17973
R:He, J.S.; Ruettlinger, R.T.; Liu, H.M.; Fulco, A.J.
Biochim. Biophys. Acta 1009, 301-303, 1989
A:Title: Molecular cloning, coding nucleotides and the deduced amino acid sequence of P
A:Reference number: S07764; MUID:90089408
A:Accession: S07764
A:Molecule type: DNA
A:Residues: 1-410 <HEJ1>
A:Cross-references: EMBL:X16610; NID:939626; PIDN:CAA34612.1; PID:g39627
A:Accession: S17973
A:Molecule type: protein
A:Residues: 1-25 <HEJ2>

```

C:Genetics: A:Gene: CYP106  
C:Superfamily: Bacillus cytochrome P450 CYP106: cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase  
F:1\_410/Product: cytochrome P450 106 #status experimental <MAT>  
F:241-378/domain: cytochrome P450 homology <CYP>  
F:356/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	15.7%;	Score 341.5;	DB 1;	Length 410;
Best Local Similarity	28.0%;	Pred. No. 1.3e-18;		
Matches 97;	Conservative 57;	Mismatches 169;	Indels 23;	Gaps 7;

QY	63	WIATROGLIEAYDYUHNHESSE---CPPIREGAEADFP-----TSMDEPBOCRPALA	113
Db	45	WNVQYEHVAKQVLSNYDFESSDGRTHIIVGDANSKKKSTSPINTNLNDRPDHKKRSLL	104
QY	116	NOVVGMYVVDKLEKRIODELACSLIESLRPOGOCNFTDYAERPRTFMLAGLPREDIP	175
Db	105	AAAFTRPSLKNWBERIOAIADIEAATOKNSTINIYDSSPRESLIADLFVPAKDRY	164
QY	176	HLKYLTPOMTRPDSMFAEAKE-----ALYDILPIEQROROKPGTAISIVANGOV	228
Db	165	QFKKMWDLRFQRDYQERLEIEIQOKRAGAEYQITLPIYIEKRSNLSDDIISDLQAEV	224
QY	229	NGRITSDEAKRMGGLLVGGSLDYVNF-L-SFSMEFLAKSPYKROVELIERPELLAPACE	287
Db	225	DGETFEDEIYHATMLLGLAGVETTSHAINMFYSFLYDKRSYSLRYNNRELLAPRAVEE	284
QY	288	LLR-FSLVADGRILTSYDEPHGVOLKSGOOLIPOMISGLDERKNAFCYNDVFSR-QKY	345
Db	285	MLRYRPHISRDRPTVTKODNELIGVKLKGGDVUAIAMASCMNDEMFEINPRESDVINRPTUK	344
QY	346	SHTFEHGSHLCLGONLAREEIIYTLKEMYLTR-----IPDESIAIP	385
Db	345	KHLTFGMRPFLCGLAPLAREMKIITLAPLAEASHPERPEDELEP	390

```

RESULT 11
B42606
cytochrome P450 CYP1B1 - Saccharopolyspora erythraea
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Saccharopolyspora erythraea
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C:Accession: B42606
R:Andersen, J.F.; Hutchinson, C.R.
J. Bacteriol. 174, 725-735, 1992
A:Title: Characterization of Saccharopolyspora erythraea cytochrome P-450 genes and
A:Reference number: A42606; MUID:92121109
A:Accession: B42606
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA; protein
A:Residues: 1-405 <AND>
A:Cross-references: GB:M83110; NID:g152682; PION:AAA26483.1; PID:g152684
A:Experimental source: NRR42338
A:Note: sequence extracted from NCBI backbone (NCBIP:77484)
C:Genetics:
A:Gene: CYP107B1
C:Species: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:238-374/domain: cytochrome P450 homology <CYP>
F:352/binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match      15.6%  Score 339.5;  DB 1;  Length 405;
Best Local Similarity 25.8%;  Pred. No. 1.8e-18;
Matches 108;  Conservative 74;  Mismatches 183;  Indels 53;  Gaps 10;

QY      18  VPEHLVFD----FDMYNSNLSAGVQEAAMVAYLQESNVDLWTRCGN-GHWIATRGLIR 72
      ||: || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      6  VPDLLAFQDRAVQGRHN-----RYAKMREEPVQRI--RTVNGIDAMLIRIYEYK 53

Y       73  EAYEDYRHFSSECPFIPREAGEAYDELPTSM-----DPEQQRQFR 112

```

Db 54 QALID-----PRLAKDGRTOQIIEKRLADAERRRPGSPDLGPHMLTDPDHTRLR 105  
113 ALANOVGMPPVVDKLENIOLACSLIESLRPOGOCNFTEDYAEPPRIRIFMLAGPEE 172  
Db 106 KLVKAKFARARVGLRPRIOITDDLDRLAGSEVDLIDEPFPLITYISELMGEVDS 165  
173 DIPHLKYLTDQMTRPDGSMTFAEAK--ALYDLIPRIEGRORPGTDAISIANGOVNG 230  
166 RRDPSRMTNVLV--DQSGEQAQAQASVAMVETLIELAKRTEPGDGLLTALLEAVEGD 223  
231 RPTSPKARMCGLLYVGLDVTYVNFSLSMELASPEHROELIERPELIPACEELR 250  
224 DRISEGELIAMVFLLVAGHETVNLGNCVLSLGNPDGLALRNDPDLPLPAIEETLR 283  
291 RFLVADG--RLITSDYEFHGVOLKKGDQILLPOMLSGLDERKNACPMHVDPSKOVSH 348  
284 YESPVANGTFRHRAEAVREDVDYIPGELVWVALGAAANRGEFEDPDRTDITRETTGHV 343  
Db 349 TFGHSHLCLGOHLARREIIVTLKEMLTRIPDSIA--PGAQIQHKSIGVQVALPL 404  
344 AFCHGHFPCGALALRLAEQIANGRLERPRDLRMASPDGLMWRPSVLMRGLEKLPLV 401

## RESULT 12

E69611  
cytochrome P450 CYP4 - Bacillus subtilis  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Bacillus subtilis  
C.Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 18-Feb-2000  
C.Accession: E69611; 144774  
R.Kunat, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berlet  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fultz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallier  
Iech, J.; Hamwood, C.R.; Henault, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
A:Authors: Schlaich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsitra, P.; Tognoni, A.; Tosato, K.; Uchiyama,  
T.; Whittes, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunori, A.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumschein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033  
A:Accession: E69611  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-410 <KUN>  
A:Cross-references: GB:299117; GB:AL009126; NID:g2634966; PIDN:CAB14615.1; PID:el183903;  
A:Experimental source: strain 168  
R.Belitsky, B.R.; Gustafsson, M.C.U.; Sonenshein, A.L.; von Wachenfeldt, C.  
J. Bacteriol. 179, 5448-5457, 1997  
A:Title: An lrp-like gene of Bacillus subtilis involved in branched-chain amino acid tra  
A:Reference number: Z22837  
A:Accession: T44774  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-410 <BEL>  
A:Cross-references: EMBL:Y11043; PIDN:CAV1937.1  
A:Experimental source: strain 1A1  
C:Genetics:  
A:Gene: cypA  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase  
F:245-381/Domain: cytochrome P450 homology <CYP>

Query Match 15.6%; Score 339; DB 1; Length 410;  
Best Local Similarity 27.5%; Pred. No. 2e-18;  
Matches 87; Conservative 69; Mismatches 138; Indels 22; Gaps 5;

QY 104 DPEOROFRALANOVGMPPVVDKLENIOLACSLIESLRPOGOCNFTEDYAEPPRIR 163  
113 ALANOVGMPPVVDKLENIOLACSLIESLRPOGOCNFTEDYAEPPRIRIFMLAGPEE 172  
Db 100 DPEDNHRLRLTVQAKTTHNIILOLEKIGHIADSLDKQPNKFMVLDYAPRLPIYI 159  
QY 164 MLAGPEEDIPHLKYLTDQMTRPDGSMTFAEAKL-----YDLIPRIEGROR 213  
160 SEMIGIPLEDROKFRVWSQA-----IDFSDAERLQENDHLLGEFEVLESLVKRRR 213  
QY 214 KPGTDVSIYANGOVGRPTSDKARMCGLLYVGLDVTYVNFSLSMELASPEHROE 273  
214 EPAGDLISALIQAESGTOLSTELYSIMLLTVAGHETVNLITMYALMCHNDQLEK 273  
QY 274 LIERPELIPACEELLRPSLV--ADGRILTSDFEFGVOLKKGDQILLPOMLSGLDER 331  
274 LRQPDLMNSAIEALRPHSPVELTIRWTAEPFLLHGOEIKKKDVIISLSANDEKI 333  
Db 332 NACPMHVDPSROKVSHTTFGHSHLCLGOHLARREIIVTLKEMLTRIPDSIA--PGAQIQH 391  
334 FPNADIFDIERKNRHRIRAFCHGNHFCIGQALRLAEKAIISTLLRRCPNITQL-KGEKKOM 392  
QY 392 K---SGIVSQVALPL 404  
393 KWKGNFLMRALIELPLI 408

## RESULT 13

A35401  
cytochrome P450 105A1 - Streptomyces griseolus  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Streptomyces griseolus  
C.Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 04-Mar-2000  
C.Accession: A35401  
R.Omer, C.A.; Lenstra, R.; Little, P.J.; Dean, C.; Tepperman, J.M.; Leto, K.J.; Romess  
J. Bacteriol. 172, 3335-3345, 1990  
A:Title: Genes for two herbicide-inducible cytochromes P-450 from Streptomyces griseo  
A:Reference number: A35401; MUID:90264332  
A:Accession: A35401  
A:Molecule type: DNA  
A:Residues: 1-406 <OME>  
A:Cross-references: GB:M32238; NID:9153477; PIDN:AAA26823.1; PID:9153478; GB:M36480  
C:Genetics:  
A:Gene: CYP105A1  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase  
F:241-377/Domain: cytochrome P450 homology <P45>  
F:355/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 15.4%; Score 336; DB 2; Length 406;  
Best Local Similarity 26.6%; Pred. No. 3.3e-18;  
Matches 105; Conservative 74; Mismatches 192; Indels 24; Gaps 11;

QY 31 PSLMSAGVQ--EAMAVLQESNPVDLWTRCNGH-WIATRGQILRAYEDYR----- 79  
113 ALANOVGMPPVVDKLENIOLACSLIESLRPOGOCNFTEDYAEPPRIRIFMLAGPEE 172  
Db 17 PNRSRCPYQLDLPDYALDRLPGLHVTLYDGRQAVVWVNHKARILDLDPRLSSNRD 76  
80 HPSSECPPLP--REAGEAVDFTPTSMDDPEQGRFRALANOVGMPPVVDKLENIOLAC 137  
77 NPFAITSPREAVRESFOA-----IGDPPHGTGRKMTISEFVKKIKKRPVEVEVNG 132  
QY 138 LIESLRPOG--QCNTEDYAEPPRIRIFMLAGLPEEDIPHLKYLTDQMTRPDGSMTFAE 196  
133 FIDEMLAGPTADLVQFALPVPVSWVICRLGVYADHDFQDASRLVOSTDAGSALNA 192  
Db 197 KEALYLLIPRIEGRORPGTDAI--SIYANGOVGRPTSDKARMCGLLYVGLDVTYV 255  
193 RNDLAGYLDGLITQFQTEBAGLVGALVADQLNG--IDREELISTAMKLILVGHETTAS 251  
QY 256 FLFSMEPLAKSPENHROELIERPELIPACEELLRPSLV--VADGRILTSDFEFGVOLK 313  
252 MTSLSVITLLDHPEDYALRADRSIVGAVEELRLATADTAGGVATADIEVSHLR 311



...

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 13:04:52 ; Search time 27.02 Seconds

(without alignments)  
474.815 Million cell updates/sec

Title: US-09-246-451-12

Perfect score: 2179

Sequence: 1 TTERIQSNANLAPLEPPHYE.....IVSGVQALPLVWDPAFTRKAV 414

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2169	99.5	414	1 CPXA_PSEPU	P00183 pseudomonas
2	584.5	26.8	414	1 Y4VG_RHISN	053215 rhizobium s
3	378	17.3	410	1 CPXM_BACME	006069 bacillus me
4	374	17.2	393	1 YH85_MYCTU	053936 mycobacteri
5	372	17.1	396	1 Y71B_BACSU	034374 bacillus su
6	366	16.8	398	1 YZ1B_MYCTU	053563 mycobacteri
7	358.5	16.5	405	1 CPXM_BACSU	P27632 bacillus su
8	346	15.9	402	1 CPXE_STRGO	P18327 streptomyc
9	342.5	15.7	428	1 YM6E_MYCTU	050696 mycobacteri
10	341.5	15.7	410	1 CPXI_SACME	P14762 bacillus me
11	339.5	15.6	405	1 CPXM_SACER	P33271 saccharopol
12	339	15.6	410	1 CPXI_BACSU	008469 bacillus su
13	336	15.4	405	1 CPXE_STRGO	P18326 streptomyc
14	324	14.9	428	1 CPXL_PSESP	P33006 pseudomonas
15	322.5	14.6	395	1 BI0I_BACSU	P53554 bacillus su
16	318.5	14.6	399	1 FAS1_RHOFA	P46373 rhodococcus
17	317.5	14.6	405	1 YC5E_MYCTU	011062 mycobacteri
18	315	14.5	412	1 CPXH_STRGR	P26911 streptomyc
19	310	14.2	402	1 Y180_MYCTU	008464 mycobacteri
20	309	14.2	408	1 NOR_FUSOX	P23295 fusarium ox
21	302.5	13.9	408	1 NOR2_CYLTO	012599 cylindrocac
22	301	13.8	381	1 CPXG_STRSO	P23296 streptomyc
23	294.5	13.5	403	1 NOR1_CYLTO	000616 cylindrocac
24	291.5	13.4	433	1 YZ4S_MYCTU	P71836 mycobacteri
25	288.5	13.2	404	1 CPXJ_SACER	Q00441 saccharopol
26	271	12.4	400	1 CPXE_RHISN	P55544 rhizobium s
27	271	12.4	414	1 Y778_MYCTU	P77903 mycobacteri
28	270.5	12.4	397	1 CPXO_SACER	P48635 saccharopol
29	259	11.9	401	1 CPXE_BRARA	059203 bradyrhizob
30	240	11.0	396	1 YW7E_MYCTU	059571 mycobacteri
31	231	10.6	436	1 THCB_RHOER	P43492 rhodococcus
32	229	10.5	402	1 Y76E_MYCTU	P77902 mycobacteri
33	218	10.0	422	1 CPXC_AGRTE	P24466 agrobacteri

# ALIGNMENTS

RESULT	ID	CPXA_PSEPU	STANDARD	PRT	414 AA.
AC	P00183:				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	13-AUG-1987 (Rel. 05, Last sequence update)				
DT	15-FEB-2000 (Rel. 39, Last annotation update)				
DE	CYTCHROME P450-CAM (EC 1.14.15.1) (CAMPOR 5-MONOOXYGENASE).				
GN	CAMC OR CYP101.				
OS	Pseudomonas putida.				
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;				
OC	Pseudomonas.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 86111751.				
RA	Unger B.P., Gunsalus I.C., Slinger S.G.:				008362 mycobacteri
RT	"Nucleotide sequence of the Pseudomonas putida cytochrome P-450cam				059572 mycobacteri
RT	gene and its expression in Escherichia coli."				059204 bradyrhizob
RL	J. Biol. Chem. 261:1158-1163(1986).				034926 bacillus su
RN	[2]				
RP	SEQUENCE OF 385-414 FROM N.A.				
RC	STRAIN-ATCC 17453:				
RX	MEDLINE: 90130389.				
RA	Koga H., Yamaguchi E., Matsunaga K., Aramaki H., Horituchi T.:				P55343 rhizobium s
RT	"Cloning and nucleotide sequences of NADH-putidaredoxin reductase				055080 sulfolobus
RT	gene (camA) and putidaredoxin gene (camB) involved in cytochrome				002318 h cytochrom
RT	P-450cam hydroxylase of Pseudomonas putida."				P14779 bacillus me
RL	J. Biochem. 106:831-836(1989).				029496 ovis aries
RN	[3]				
RP	SEQUENCE.				
RX	MEDLINE: 83030788.				
RA	Haniu M., Armes L.G., Yasunobu K.T., Shastri B.A., Gunsalus I.C.:				P24467 agrobacteri
RT	"Amino acid sequence of the Pseudomonas putida cytochrome P-450. II.				002390 sus scrofa
RT	Cyanogen bromide peptides, acid cleavage peptides, and the complete				033180 mycobacteri
RT	sequence."				
RL	J. Biol. Chem. 257:12664-12671(1982).				
RN	[4]				
RP	X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).				
RX	MEDLINE: 86143817.				
RA	Fitzel B.C., Weber P.C., Hardman K.D., Saleme F.R.:				
RT	"Structure of ferricytochrome c' from Rhodospirillum rubrum at				
RT	1.67-A resolution."				
RL	J. Mol. Biol. 186:627-643(1985).				
RN	[5]				
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).				
RX	MEDLINE: 86059514.				
RA	Poulos T.L., Fitzel B.C., Gunsalus I.C., Wagner G.C., Kraut J.:				
RT	"The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450."				
RL	J. Biol. Chem. 260:16122-16130(1985).				
RN	[6]				
RP	X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).				
RX	MEDLINE: 98019009.				
RA	Schlichting I., Jung C., Schulze H.:				
RT	"Crystal structure of cytochrome P-450cam complexed with the (1S)-				
RT	camphor enantiomer."				
RL	FEBS Lett. 415:253-257(1997).				

RN [7]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RX MEDLINE: 98313255.  
RA Vidakovic M., Sligar S.G., Li R., Poulos T.L.;  
RT "Understanding the role of the essential Asp51 in cytochrome P450cam  
RT using site-directed mutagenesis, crystallography, and kinetic solvent  
RT isotope effect.";   
RL Biochemistry 37:9211-9219(1998).  
RN [8]  
RP STRUCTURE BY NMR.  
RX MEDLINE: 97459726.  
RA Moura C., Bondon A., Simmoneaux G., Jung C.;  
RT "1H-NMR study of diamagnetic cytochrome P450cam: assignment of heme  
RT resonances and substrate dependence of one cysteine beta proton.";   
RL FEBS Lett. 414:203-208(1997).  
CC -1- FUNCTION: INVOLVED IN A CAMPHOR OXIDATION SYSTEM.  
CC -1- CATALYTIC ACTIVITY: CAMPHOR + PUTIDAREDOXIN + O(2) = 5-EXO-  
CC HYDROXYCAMPHOR + OXIDIZED PUTIDAREDOXIN + H(2)O.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M12546; AAA25760.1; -;  
CC EMBL: D00528; BAA00412.1; -;  
DR PIR; A25660; Q4PSCP.  
DR PDB; 2CPE; 15-APR-91.  
DR PDB; 3CPE; 15-APR-91.  
DR PDB; 4CPE; 15-JUL-91.  
DR PDB; 5CPE; 15-JUL-91.  
DR PDB; 6CPE; 15-JUL-91.  
DR PDB; 7CPE; 15-JUL-91.  
DR PDB; 8CPE; 15-JUL-91.  
DR PDB; 1CP4; 15-JAN-93.  
DR PDB; 2CP4; 15-JAN-93.  
DR PDB; 3CP4; 15-JAN-93.  
DR PDB; 4CP4; 15-JAN-93.  
DR PDB; 5CP4; 16-SEP-98.  
DR PDB; 6CP4; 16-SEP-98.  
DR PDB; 1NOO; 08-MAR-96.  
DR PDB; 1PHA; 31-OCT-93.  
DR PDB; 1PHB; 31-OCT-93.  
DR PDB; 1PHC; 31-OCT-93.  
DR PDB; 1PHD; 31-OCT-93.  
DR PDB; 1PHE; 31-OCT-93.  
DR PDB; 1PHF; 31-OCT-93.  
DR PDB; 1PHG; 31-OCT-93.  
DR PDB; 1AKD; 19-NOV-97.  
DR PIRAM: PF00067; P450; 1.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW 3D-structure.  
FT INIT\_MET 0  
FT BINDING 357  
FT CONFLICT 55  
FT CONFLICT 276  
FT CONFLICT 361  
FT CONFLICT 407  
FT CONFLICT 407  
FT TURN 16  
FT HELIX 20  
FT STRAND 23  
FT TURN 28  
FT TURN 32  
FT HELIX 34  
FT HELIX 38  
FT TURN 48  
FT STRAND 53  
FT HELIX 58

FT	STRAND	62	65
FT	HELIX	68	76
FT	TURN	78	80
FT	STRAND	81	82
FT	TURN	83	84
FT	HELIX	90	95
FT	TURN	99	102
FT	TURN	105	107
FT	HELIX	108	119
FT	HELIX	121	145
FT	TURN	146	146
FT	STRAND	147	149
FT	HELIX	150	153
FT	TURN	154	156
FT	HELIX	157	167
FT	TURN	168	168
FT	HELIX	171	185
FT	HELIX	193	213
FT	TURN	219	224
FT	TURN	225	225
FT	STRAND	227	228
FT	TURN	229	230
FT	STRAND	231	232
FT	HELIX	235	250
FT	TURN	251	252
FT	HELIX	253	266
FT	HELIX	268	276
FT	HELIX	278	291
FT	TURN	292	292
FT	STRAND	295	295
FT	STRAND	297	301
FT	STRAND	305	307
FT	TURN	308	309
FT	STRAND	310	312
FT	TURN	314	315
FT	STRAND	317	320
FT	HELIX	322	324
FT	TURN	325	327
FT	TURN	329	331
FT	TURN	335	336
FT	TURN	340	341
FT	TURN	349	350
FT	HELIX	353	355
FT	TURN	358	359
FT	HELIX	360	377
FT	STRAND	382	383
FT	TURN	385	386
FT	STRAND	391	392
FT	STRAND	396	396
FT	STRAND	398	399
FT	STRAND	403	405
FT	HELIX	408	410
SO	SEQUENCE	414 AA;	46538 MW; ECA610293A9D6207 CRC64;

Query Match 99.5%; Score 2169; DB 1; Length 414;  
Best Local Similarity 99.5%; Pred. No. 2.6e-158;  
Matches 412; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1	TTETIOSNANLAPLPHVPEHLVFDMDYNPNSLSAGVQENAVLQESNVPDLVWTRCNG	60
DB	1	TTETIOSNANLAPLPHVPEHLVFDMDYNPNSLSAGVQENAVLQESNVPDLVWTRCNG	60
QY	61	GHWIAIRGOLIRAYEDYRHSSECPPIREAGEAYDFTSMDDPEQORFALANQVYG	120
DB	61	GHWIAIRGOLIRAYEDYRHSSECPPIREAGEAYDFTSMDDPEQORFALANQVYG	120
QY	121	MPVVDKLENIQELACSLIESLRPOGQCNFTEDYAPFPPIRIMLLAGLPEEDIPIHLKYL	180
DB	121	MPVVDKLENIQELACSLIESLRPOGQCNFTEDYAPFPPIRIMLLAGLPEEDIPIHLKYL	180
QY	181	TDQMTRPDGSMTFAEAKKALYDYLPIIEQRQKPGTDAISIVANGQVNGRPITSDENAKR	240

Db 181 TDQMRPQSGMTEFAAKALDYLLIPITIEQRQKPGDTAISIVANGQVNGRPTSDCAK 240

Qy 241 MCGLLVAGCLDTVAVNFSLFSMEFLAKSPENHQELIERELLIPACEELLRRSLVADGRI 300

Db 241 MCGLLVAGCLDTVAVNFSLFSMEFLAKSPENHQELIERERIPAAEEELLRRSLVADGRI 300

Qy 301 LTSDFEFHGVOLKCKDQDILLPOMLSGLDERKACPMHNDPFSROKSHHTFGSHLCLGQ 360

Db 301 LTSDFEFHGVOLKCKDQDILLPOMLSGLDERKACPMHNDPFSROKSHHTFGSHLCLGQ 360

Qy 361 HLAERREIVTLKEMLTIRIPDFSIAFGAIOIHKSGIVSGVALPLVMDPATTKAV 414

Db 361 HLAERREIVTLKEMLTIRIPDFSIAFGAIOIHKSGIVSGVALPLVMDPATTKAV 414

RESULT 2

Y4VG\_RHISN ID Y4VG\_RHISN STANDARD; PRT: 414 AA.

AC 053215;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE PROBABLE CYTOCHROME P450 Y4VG (EC 1.14.14.-).

GN Y4VG.

OS Rhizobium sp. (strain NGR234).

OC Plasmid sym PNGR234e.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.

RN 11

RP SEQUENCE FROM N.A.

RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A., Perret X.;

RT "Molecular basis of symbiosis between Rhizobium and legumes."; Nature 387:394-401(1997).

RL 12

RP SEQUENCE FROM N.A.

RX MEDLINE: 963899014.

RA Freiberg C., Perret X., Broughton W.J., Rosenthal A.;

RT "Sequencing the 500-kb GC-rich symbiotic replicon of Rhizobium sp. NGR234 using dye terminators and a thermostable 'sequenase': a beginning.";

RT Genome Res. 6:590-600(1996).

CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

CC CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC CC

DR EMBL: Z68203; CAAG2422.1; -

DR EMBL: AE000101; AAB91895.1; -

DR HSSP: P00183; 3CP4.

DR PFAM: PF00067; P450. 2.

DR PROSITE: PS00086; CYTOCHROME\_P450; FALSE NEG.

KW Hypothetical protein; Oxidoreductase; Monooxygenase;

KW Electron transport; Heme; Plasmid.

FT BINDING 364 HEME (BY SIMILARITY).

SO SEQUENCE 414 AA: 45810 MW: 888598E99315BB84 CQC64;

Query Match 26.8% Score 584.5; DB 1; Length 414;  
 Best Local Similarity 33.8%; Pred. No. 1.9e-37;  
 Matches 141; Conservative 65; Pmiatons 172; Indels 39; Gaps 9;  
 QY 13 PLPLPVPPELVFEDDMYPSNLSA---GVQAEANVLQESNPVDLW----TRCNGHMTA 65  
 1:1 111 11 1:1 : : 1 : : : : 11 1:1

```

Db 13 PIPDHVPALVRHHSLSFTSPGMAPTNPNDPAHAACVADDOCPRIFFYSPSNRDRGTYVI 72
Qy 66 TRGQLIREAYDYRRHFSECEPIPREACEAYDIPTSMDPEQROFRALANOVGMYPVD 125
Db 73 TRARQORFVLEDTETFESSHRSIFASALCEHNPVPIPLEDDPAHGVALFNLPLFSSRYL 132
Qy 126 KLENRIOELASLIESL-RPGSGCNFTEDVAPPIRIETMLLAIGE-----EDI 174
Db 133 ALEPTIHARACALIDCIKAKETSCVMDFALPFFSVFLSGISORSEVLVGMWSDL 192
Qy 175 PHLKYLTDQMTRPDQSMTEFAKEALYLYLPIIEQ---BRQKPGTAPAIIVANGVNG 230
Db 193 LH-----GN---AEKRRAAASVAFIDEMAAAMRKSPAVDFMTFYVQAKIEG 237
Qy 231 RPIITSEAKRMCGLLVGLDVTYVNFSLFSMEFLAKSPENROELIERPELLPAACEELLR 290
Db 238 RSLTEEEVRGIGVLEFVAGLDTVAALIGFDMAVLAIRNKHQELLRNBPARGLAEEELLR 297
Qy 291 RPSLVAQORILTSDFEFVGLKKGDDLLPQMLSGLDERKNACPMHVDSEKQVSHTFE 350
Db 298 AVSTVQIIRVATKIDLEFGVPIDREDVYSCPMIANRPSSEFCPTDLARODQHTAF 357
Qy 351 GGHSGHLCGHLARREIIVTLKEMLTIRPDPSIAPG-NOIQHKSIGVGYQALPLW 406
Db 358 GYCPHLCHGAILRAREIYIGLEMLAIRIPARIKEGTAPITH-GGHVFGISNILLTW 413

RESULT 3
CPXM_BACME STANDARD: PRT: 410 AA.
AC 006069;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-FEB-2000 (Rel. 39, last annotation update)
DE CYCROCHROME P450(MEG) (EC 1.14.99.-) (STEROID 15-BETA-HYDROXYLASE)
DE (STEROID 15-BETA-MONOOXYGENASE).
GN CYP106A2.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN-ATCC 13368;
RC MEDLINE: 94049677.
RA Rauschenbach R., Isernhagen M., Noeske-Jungblut C., Boold W.,
RA Siewert G.;
RT "Cloning, sequencing and expression of the gene for cytochrome
RT P450meg, the steroid-15 beta-monooxygenase from Bacillus megaterium
RT ATCC 13368.";
RL Mol. Gen. Genet. 241:170-176(1993).
RN [2]
RC CHARACTERIZATION.
RC STRAIN-ATCC 13368;
RC MEDLINE: 79194051.
RA Berg A., Ingelman-Sundberg M., Gustafsson M.;
RT "Purification and characterization of cytochrome P-450meg.";
RL J. Biol. Chem. 254:5264-5271(1979).
RN [3]
RC CHARACTERIZATION.
RC STRAIN-ATCC 13368;
RC MEDLINE: 82091079.
RA Berg A., Raftter J.J.;
RT "Studies on the substrate specificity and inducibility of cytochrome
RT P-450meg.";
RL Biochem. J. 196:781-786(1981).
CC -1- FUNCTION: HAS THE CAPACITY TO HYDROXYLATE CERTAIN STEROIDS IN THE
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

DT 01-FEB-1995 (Rel. 31, created)  
 DT 01-FEB-1995 (Rel. 31, last sequence update)  
 DT 15-FEB-2000 (Rel. 39, last annotation update)  
 DE CYTOCHROME P450(MEG) (EC 1.14.99.-) (STEROID 15-BETA-HYDROXYLASE)  
 DE (STEROID 15-BETA-MONOOXYGENASE).  
 GN CYP106A2.  
 OS Bacillus megaterium.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN-ATCC 13368;  
 RX MEDLINE; 94049677.  
 RA Rauschenbach R., Isernhagen M., Noeske-Jungblut C., Boiold W.,  
 RA Stewert G.;  
 RT "Cloning sequencing and expression of the gene for cytochrome  
 RT P450meg, the steroid-15 beta-monooxygenase from Bacillus megaterium  
 RT ATCC 13368.";  
 RL Mol. Gen. Genet. 241:170-176(1993).  
 RN [2]  
 RP CHARACTERIZATION.  
 RP STRAIN-ATCC 13368;  
 RX MEDLINE; 79194051.  
 RA Berg A., Ingelman-Sundberg M., Gustafsson M.;  
 RT "Purification and characterization of cytochrome P-450meg.";  
 RL J. Biol. Chem. 254:5264-5271(1979).  
 RN [3]  
 RP CHARACTERIZATION.  
 RP STRAIN-ATCC 13368;  
 RX MEDLINE; 82091079.  
 RA Berg A., Rafters J.J.;  
 RT "Studies on the substrate specificity and inducibility of cytochrome  
 RT P-450meg.";  
 RL Biochem. J. 196:781-786(1981).  
 CC -1- FUNCTION: HAS THE CAPACITY TO HYDROXYLATE CERTAIN STEROIDS IN THE  
 CC 15-BETA POSITION.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way





RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF015825; AAC46317.1; -  
DR EMBL; Z99110; CAB13078.1; -  
DR SUBLIST; BG13195; YUIB.  
DR PFAM; PF00067; P450; 1.  
DR PRINTS; PR00359; BP450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
FT BINDING 349 349 HEME (BY SIMILARITY).  
SQ SEQUENCE 396 AA; 44990 MW; 9A89CF12613DBCFB CRC64;  
  
Query Match 17.1%; Score 372; DB 1; Length 396;  
Best Local Similarity 30.7%; Pred. No. 3e-21;  
Matches 101; Conservative 54; Mismatches 156; Indels 18; Gaps 7;  
  
OY 71 IREAYEDYRHHSSCPPIPREAGEAYDFIPTSMDPPQORORALANOVGVYDKLENR 130  
Db 56 VKKVGKDELFS-C-MPOQTSSIGNSI-IMDDPKHTKIRSVYKAFTRVMKQWEPK 111  
OY 131 IOELACSLIESLRQGGCNFTEDYAEPPIRIFMLAGLPREDIPHLKYLTDMT----- 185  
Db 112 IOETDELIOKFGSSEDFLVHDSYPLPVIVISLIGVPSAHMQGFAMSDLLVSTPKD 171  
OY 186 -RPGSMFPAE-----AKALYDYLPIIEORRQKPTDAISIVANGVNGRPITSDKKR 240  
Db 172 KSEAEKAFLEERDKCEBELAFAFGIIEEKRKNPEODIISILVEATEGKLSGEELIP 231  
OY 241 MCGLLVAGGLTVVNFSLFSMEPLAKSPENHOLELERPELIPACEELLR-RFSLVADGR 299  
Db 232 FCTLLLVANGNETTNNLSNAMYSLIETPGVTEELRSHELMPOAVEELRRAPAPVLR 291  
OY 300 ILTSDYEFHGVOLKGDQILLPOMLSGLDERKNACPMHVDERSQKVSHTTGHSGSLCLG 359  
Db 292 IAKRTELGHLIKEDGVNLAFAVASANDKAFDRPHMFDLRHNPRIATGHHGHPCLG 351  
OY 360 CHLAREELIVLKEMLRIPDF---STAP 385  
Db 352 APLARLEANIALTSLISAFPHMECVSTIP 380  
  
RESULT 6  
Y218\_MYCTU STANDARD; PRT; 398 AA.  
ID Y218\_MYCTU  
AC 053563;  
DT 15-FEB-2000 (Rel. 39, Created)  
DT 15-FEB-2000 (Rel. 39, Last sequence update)  
DE 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PUTATIVE CYTOCHROME P450 RV3518C (EC 1.14.-.-).  
GN RV3518C OR MTVO23.25C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
MEDLINE; 98295987.  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,  
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RP "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AL022022; CAA17755.1; -  
DR TUBERCULIST; RV3518C; -  
DR PFAM; PF00067; P450; 1.  
DR PRINTS; PR00359; BP450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
FT BINDING 340 340 HEME (BY SIMILARITY).  
SQ SEQUENCE 398 AA; 44398 MW; BCF3C23BC85767F CRC64;  
  
Query Match 16.8%; Score 366; DB 1; Length 398;  
Best Local Similarity 28.5%; Pred. No. 8.6e-21;  
Matches 108; Conservative 69; Mismatches 168; Indels 34; Gaps 8;  
  
OY 55 WTRCN-----GGHMTATRGQLIREAYEDYRHHSS-----ECPEIIPREAGEAYDFIPT 101  
Db 26 WMRAQPVYFRDRNGLAASSTYQAVIDAERPELSNAGIRPDQALP-----MMI 76  
OY 102 SMDPEQORORALANOVGVYDKLENRIEOLACSLIESLRQGGCNFTEDYAEPPIR 161  
Db 77 DMDPEHLLRRKLVNAGFTRRKRVKDEKSIALCDTLLDAVCERQCDPVDLAAPLPM 136  
OY 162 IFMLLAGL-PEEDIPLHLK-----YLTDMTPRPGSMFPAEALYDYLPIIEORQ 213  
Db 137 VIGDMGVPRQDRMFLMSDDLVTFLSSHYVSOEDPQITM-DARAAYUDFTRATIAARR 195  
OY 214 KPGTDAISIVANGVNGRPITSDKARKGGLLVGGLTVVNFSLFSMEPLAKSPENHROE 273  
Db 196 DPTDVLVSLSSEVDGRSLSDDELVMETLLILGGDETTHNTLSGTEQLLRNDQMDL 255  
OY 274 LIERPELIPACEELLRFSLVAD-GRILTSDYEFHGVOLKGDQILLPOMLSGLDERKN 332  
Db 256 LQRPSSLPGALIEMLRWTAIVKKNKCRVLTADTEFHGATLCAGKEMLLFESANFDEAVF 315  
OY 333 ACPMHVDSRQKVSHTTGHSGSLCLGHLARRELIIVLKEMLRIPDFSTAPGQI--Q 390  
Db 316 CEPKFDVQRRPNSHLAGFTGTHFCGLQNLARLELSTMTERTVLRRLVADSDVLR 375  
OY 391 HKSGIVSGVALPLVMDPA 409  
Db 376 RPAENVSGLESMPTVFTPS 394  
  
RESULT 7  
CPXM\_BACSU STANDARD; PRT; 405 AA.  
ID CPXM\_BACSU  
AC P27632;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE CYTOCHROME P450 109 (EC 1.14.-.-) (ORF405).  
GN CYP109.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=W23;

```
RX MEDLINE; 91192601.
RA Ahn R.S., Wake R.G.;
RT "Variations and coding features of the sequence spanning the
RT replication termini of Bacillus subtilis 168 and W22 chromosomes.";
RL Gene 98:107-112(1991).
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24523; AAA22720.1; -.
DR HSSP; Q00441; IOXA.
DR PRAM; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Heme.
FT BINDING 351 351 HEME (BY SIMILARITY).
SQ SEQUENCE 405 AA; 45845 MW; 133D5BEA7AE3C04 CRC64;

Query Match 16.5%; Score 358.5; DB 1; Length 405;
Best Local Similarity 25.2%; Pred. No. 3.3e-20;
Matches 113; Conservative 72; Mismatches 169; Indels 95; Gaps 10;

QY 1 TTTETQSN-----ANLAPLP-HYPEHLVDFEDMNPNSNLGAVY-----QENAAVLQESN 49
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 2 TNGPARSSKKRRYANLIPMELHSEKDRLEFPPIYDKRRRSPRYRDPDRCMQVFK--- 58
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 50 VPDLVWTRCNGHVIATRGOLIRAEVDEYRHSSECPPIPREAGAVDPIPTSMPPBQR 109
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 59 -----YDVQVFLKNPKLFSSKRIQIESILT-MDPKHT 92
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 110 QFRALANQVMPYVDKLENIQELACSLIESLRPGQCNTEDEYAEFFIRIFMLAGL 169
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 93 KLRALVSAFTPKAVKQLETRIKQVTAFLQEAROKSTIIDIEDFAGLPVITIIAEMGA 152
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 170 PEEDIPHLKYLTDQMTPRDGSMTFAEAKA-----LYDLIPITIEOR 211
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 153 PIEDRHILIKTYDVLV-----AGAKDSSDKAVADVWNRDRGHAFSLSYFEDILSKR 204
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 212 ROKGCTDAISLVANGQVNGRPTSDCAKRMCGLLVGLDVTVNFSLFSMEFLAKSPBHR 271
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 205 RAEPREDLMTMLQAEIDGELYTEOLIGFCILLVAGNETTMDIANAVRYLTEDSVYQ 264
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 272 QELIERPELPAACEELIRPSLV-ADGRILTSYEFHGVQKKGGDQIILPOMLSGDER 330
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 265 QQVQNTNDNVANVIEELIKRYSVQALGRVATEDTELGGVFIKRGSSVTSIASANDED 324
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 331 KNAQPMHVDSEKQVSHTTFGHSHLCLGQHLARREIIVTLK-----EML 375
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 325 KFCRPFCKIDRPSYPHLSFGFGIHFCGAPLARLEANIALLSSLSACIEKAHDEKL 384
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 376 TRIDFSLAFCAGQIQHKSIGVQDALP 404
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 385 EAIP-----SPFVGVKRLPV 400
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 8
CPXF_STRGO STANDARD; PRT; 402 AA.
AC P18327;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450-SU2 (EC 1.14.-.-) (P450-CVBI) (CYP105B1).
```

```
GN CYP105B1 OR SUBC.
OS Streptomyces griseolus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-32.
RC STRAIN-ATCC 11796;
RX MEDLINE; 90264332.
RA Omer C.A., Lenstra R., Little P.J., Dean C., Tepperman J.M.,
RA Lato R.J., Romesser J.A., O'Keefe D.P.;
RT "Genes for two herbicide-inducible cytochromes P-450 from
RT Streptomyces griseolus.";
RL J. Bacteriol. 172:3335-3345(1990).
CC -!- FUNCTION: METABOLISM OF A NUMBER OF SULFONYLUREA HERBICIDES.
CC -!- INDUCTION: BY HERBICIDE.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M32239; AAA26825.1; -.
DR PIR; B35401; B35401.
DR HSSP; P23295; 2ROM.
DR PRAM; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Heme.
FT INT MET 0
FT BINDING 351 351 HEME (BY SIMILARITY).
SQ SEQUENCE 402 AA; 44278 MW; E3B67F672C26D9D CRC64;

Query Match 15.9%; Score 346; DB 1; Length 402;
Best Local Similarity 29.2%; Pred. No. 2.9e-19;
Matches 106; Conservative 63; Mismatches 164; Indels 30; Gaps 13;

QY 63 WIATRGOLIRAEVDEYRHSSECC-----PIPREAGAVYPIPT--SMPPBQRQFRALA 115
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 51 WLVTYRHQDVRAVLGDPR-FSADARHTGFPPLTAGRGRIICITNPFLRMDDPBARLRRL 109
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 116 NOVYGMVYVDKLENIQELACSLIESLRP-OGQCNTEDEYAEFFIRIFMLAGLPEDEI 174
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 110 TADFTYAKKVEAMREPVRQLDDVDRMTTGTSADLVTERFALPLPSLVLCILGVPYEDH 169
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 175 PHL-----KYLTDQMTPRDGSMTFAEAKALYDVLPIITEOROKPGTDAIS-IVANGQVN 229
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 170 AFGERSRVLLTLINSTEPE---EVRAAQDELLEYLARLARTRKREPRDAIISRLVARELD 226
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 230 GRPTSDCAKRMCGLLVGLDVTVNFSLFSMEFLAKSPHROBELIERPELIPAAECLEL 289
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 227 DTQIAT-----MGRLLLVAAHETTANMTALSTVLLENPQRLRAEPALVAGVARELL 281
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 290 RRFSLVADG--RITTSYEFHGVQKKGGDQIILPOMLSGDERKNACR--MHVDESROXY 345
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 282 RYLTIVHNGVPRITTEVLLIGRTIAGEGVLC--MSSANNRADEVPRGGDDLDVARDAR 339
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 346 SHTTFGHSHLCLGQHLARREIIVTLKEMLTRIPDFSA-PGAQIQHKSIGI-VSQVQALP 403
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 340 RHVAFGFGVHQCGLQPLARVLEIDAIITLRLRDLALAVRHEBIPRGRMAYGVHSLP 399
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 9
YM66_MYCTU STANDARD; PRT; 428 AA.
ID YM66_MYCTU
```

AC 050696;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PUTATIVE CYTOCHROME P450 RV2266 (EC 1.14.1.1).  
GN RV2266 OR MTCY339.44C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RX MEDLINE: 98295987.  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z77163; CAB00969.1; -;  
DR HSSP: P33006; ICP.  
DR TUBERCULIST: RV2266; -;  
DR PFAM: PF00067; P450; 1.  
DR PRINTS: PR00359; BP450.  
DR PROSITE: PS00086; CYTOCHROME\_P450; FALSE\_NEG.  
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
FT BINDING 379 379 HEME (BY SIMILARITY).  
SQ SEQUENCE 428 AA; 47824 MW; 76B1F3C5AE348591 CRC64;  
Query Match 15.7%; Score 342.5; DB 1; Length 428;  
Best Local Similarity 26.3%; Pred. No. 5.9e-19;  
Matches 118; Conservative 71; Mismatches 188; Indels 69; Gaps 17;  
OY 5 IOSNANLAPRPVPHVPEHVLVDFPDWNPNSLSAGVOEAMVAIOE-----SNVPLV 54  
DB 7 IATVNGRP-PPEVP---IADIEL-----GSLDFWALDDVDVRCGAFATLRAPISF 54  
OY 55 WTR-----CNGHMWATRCQLREAYEDY-----RH-----FSS-----ECPEIPREA 92  
DB 55 WPTIELPGFVAGNGHWALTK-----YDVEFASRHDPDISSVNTITINDOTELAYF 107  
OY 93 GEATDFITSDPPEQRFRLANOVGMVYDKLENIQELACLISL---RPGGCN 149  
DB 108 GSM-----IVLDDPRHOLRSIVSRATPKVVARIEAIVRRRAHRLVSSMANNDRDAD 162  
OY 150 FTEVAAEPFPIRIFMLLAGPEEDIPHLKYLTDMT---RPGSMTFPAKEALD---Y 203  
DB 163 LVSELAGLPQIICDMNGIFKADHQRIFFHNTNYILGFGDPDLADLDFEFGQVSDIATY 222  
OY 204 LIPITEORRQKPGTDAISIVANGVNGRPITSDEAKRMCGLLVGLGDTVNFISFSMEF 263  
DB 223 ATALAEDRVNHHDLTSLVEAEVDGERSLSREIASFFILLVAGNFTETNALITHGVIA 282  
OY 264 LAKSEHNOE-LIERPELIPACCELLARFS-LVADGRILTSVDFEFGVQLKGGQIILP 321  
DB 283 LSRPEQDRMWSDFGLAPRAVEIIVWASPVVYMRRTLTODIELRGTCKMAAGDKVSLW 342

OY 322 QMLGLDERKNACPMHVDFSROKVSHTTF-GHSHLCGOLARREITVLEKWLTRIPD 380  
DB 343 YCSANROESKFPADPWTFTDLARNPNPHLFGGGAHFICGANLARREIRVAPDELRQMPD 402  
OY 381 FSIAPGAOIQRKSGIVSGVQALPLVW 406  
DB 403 V-VATEEPARLLSQFIRHGIKTLPVYW 427  
RESULT 10  
CPXL\_BACME STANDARD; PRT; 410 AA.  
ID CPXL\_BACME  
AC P14762;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DE 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME P450(BM-1) (EC 1.14.14.1).  
GN CYP106.  
OS Bacillus megaterium.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-25.  
RC STRAIN-ATCC 14581;  
RX MEDLINE: 90089408.  
RA He J.S., Ruettinger R.T., Liu H.-M., Fulco A.J.;  
RT "Molecular cloning, coding nucleotides and the deduced amino acid  
RT sequence of P-450BM-1 from Bacillus megaterium."  
RL Biochim. Biophys. Acta 1009:301-303(1989).  
RN [2]  
RP SEQUENCE OF 1-25 FROM N.A.  
RX MEDLINE: 95355495.  
RA He J.S., Liang O., Fulco A.J.;  
RT "The molecular cloning and characterization of BM1P1 and BM1P2  
RT proteins, putative positive transcription factors involved in  
RT barbiturate-mediated induction of the genes encoding cytochrome  
RT P450BM-1 of Bacillus megaterium."  
RL J. Biol. Chem. 270:18615-18625(1995).  
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED  
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X16610; CA34612.1; -;  
DR EMBL: S79230; AAC60495.1; -;  
DR PIR: S07764; OAB56M.  
DR HSSP: P33006; ICP.  
DR PFAM: PF00067; P450; 1.  
DR PRINTS: PR00359; BP450.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.  
FT BINDING 356 356 HEME.  
SQ SEQUENCE 410 AA; 47460 MW; C9AE293E76745387 CRC64;  
Query Match 15.7%; Score 341.5; DB 1; Length 410;  
Best Local Similarity 28.0%; Pred. No. 6.6e-19;  
Matches 97; Conservative 57; Mismatches 169; Indels 23; Gaps 7;  
OY 63 WIATGQLIREAYEDYRHFSSSE---CPPIRPAAGAYDFIP-----TSMDDPEQRFRLA 115  
DB 45 WNVQYEVHKOVLSTNDFSSDGGRTTFVGDNSKKKTSPTTNLTNIDPDHKKRSL 104  
OY 116 NOVGMVYDKLENIQELACLISLRRPGOCNFTEDYAEFPPIRIFMLLAGPEEDIP 175

```

Db 105 AAATPRSLAKNNEPRIKQIADLVLEIAQKNSTINIVDDLSPPRLVIADLCEVYPKDXR 164
Oy 176 HLKYLTDQMTPRDGSMTPEAKE-----ALYDLIPIIEGRORPGDAISIVANGY 228
Db 165 QPKKVVNDILFQPYDQERLEIEQEKORAGAEFYQYLPIVIEKRSMLSDIILQAEV 224
Oy 229 NGRPLTSDPAKRMGGLIYVGLDVTWNFL-SYSMETLASEPKRQILIERPELLIPACEE 287
Db 225 DQETFTDEIVATMILLGAGVETTSHTAIANMPYSFLYDKSLYSFLRNRRNLAPAAVEE 284
Oy 288 LLR-FEFLVADGRILTSDVEFEFGVOLKQDQIILLPQMLSGLERKNKACPMHYDFSR-QKV 345
Db 285 MLRYEPHLISRBRQRYTKQDQNELLGVLTKKQDDVIAMMSACNMDETMEFNFSVDIHPTUK 344
Oy 346 SHTTGHGSHLCLGQHLARREITVTLKEMLTR-----IPDFSIAIP 385
Db 345 KHLFTGNGPHFCLAGPLARLEMKIILEAFLEAFSHTEPFEDELELP 390

RESULT 11
CPKX_SACER STANDARD: PRT; 405 AA.
AC P33271;
DT 01-FEB-1994 (Rel. 28, created)
DT 01-FEB-1994 (Rel. 28, last sequence update)
DT 15-DEC-1998 (Rel. 37, last annotation update)
DE CYTOCHROME P450 107B1 (EC 1.14.-.-) (P450C11B1).
GN CYP107B1.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardiaceae; Pseudonocardiaceae;
OC Saccharopolyspora.
CC [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 172-203 AND 302-321.
RP STRAIN-NRRL 2338;
RX MEDLINE: 92121109.
RA Andersen J.F., Hutchinson C.R.;
RT "Characterization of Saccharopolyspora erythraea cytochrome P-450
RT genes and enzymes, including 6-deoxyerythronolide B hydroxylase.";
RL J. Bacteriol. 174:725-735(1992).
CC -1- FUNCTION: NOT KNOWN, PROBABLY INVOLVED IN THE CATABOLISM OF
CC OCTANE AND GAIAICOL. IT DISPLAYS A WEAK ACTIVITY IN THE O-
CC DEMETHYLATION OF 7-ETHOXYCUMARIN.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/isb-sib.ch).
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M83110; AAA26483.1; -.
DR PIR: B42606; B42606.
DR HSSP: Q00441; IOXA.
DR PFAM: PF00067; P450. 1.
DR PRINTS: PR00359; BP450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT BINDING 352 352 HEME (BT SIMILARITY).
SQ SEQUENCE 405 AA; 45238 MM; 71C93CEC1FDC53FD CRC64;

Query Match 15.6%; Score 339.5; DB 1; Length 405;
Best Local Similarity 25.8%; Pred. No. 9.2e-19;
Matches 108; Conservative 74; Mismatches 183; Indels 53; Gaps 10;

Oy 18 VPEHVPFD-----FDYVNSNLSAGVQAEAMVAVYNSPDLVWTRCNG-GHWIATRGQLR 72
Db 6 VPDLLAFPDARQQRHN-----RYARRKEEPVQRI-RTYVNGDADAMLTITREDDYK 53
Oy 73 EAYEDYRHFSESCPEPIPEAGVAYDPIETSM-----DPPEQROFR 112

```

```

Dd      54 QALD-----PRTADFGRTOOIIETKRLADAERRPESPDGLPRHMLNTDPDRPTBLR   105
Qy      113 ALANOVGMPPVDKIENRIEACLSIESLRPOGCNFTEDEYAEEPPIRIFIMLAGLPEE   172
Db      106 KLVYAFARFAARVEGRRIPRIEQITDDLLDRLAGRSVDLDIDEFAFLPTTVISELMGVSDS   165
Qy      173 DIPILKYLTDMOTPRDGSMTFAEAK--ALDYLPITEQRORRGKTDAISIVANGVNG   230
Db      166 RDDRFRSTNVLV--DGSOPEAOASVAWEVYTELIATKRTEREGDDLTLALLEAVEDG   223
Qy      231 RPITSDEAKRMCGILLVYGDLDTVVNPLSFSEMEFLAKSPENHOELERBELPAACEELLR   290
Db      224 DRLSGELIAWFLVLVAGHETYNLIGNCVSLSGNDOLAALKNDSPLLPGALEETLR   283
Qy      291 RESLVADG--RLTSDYEFHGVOALKKGDIILPOMLSGLDERKNAPMHWFSROKVSHT   348
Db      284 YESPVANTFFHTAEAVAFGDVPIPEGSLVMVALGAARNRDERFEDPRFDITRETTHV   343
Qy      349 TFGHSILCQLQHARREIIVTLKEWLTRIDPFSTA--PGAIOHKSGVISGVQALPL   404
Db      344 AFGHGHHPCVCAALARLEAQIAGVGLRLERFPDLRKMAAPDDLRFMRFSVLMRGLEXLPV   401

RESULT 12
CPXY_BACSU          STANDARD;             PRT;       410 AA.
AC      008469;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-FEB-2000 (Rel. 39, Last annotation update)
DE      CYTOCHROME P450 (EC 1.14.-.-).
GN      CYP4 OR CYP107J1.
OS      Bacillus subtilis.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group;
CC      Bacillus/Staphylococcus group; Bacillus.
[1]
RN      RN
RP      SEQUENCE FROM N.A.
RC      STRAIN=168 / BGSC1A1;
RX      MEDLINE; 974311495.
RA      Belitsky B.R., Gustafsson M.C.U., Sonenshein A.L., von Wachenfeldt C.;
RT      "An lrp-like gene of Bacillus subtilis involved in branched-chain
RL      amino acid transport."; J. Bacteriol. 179:5448-5457(1997).
[2]
RN      RN
RP      SEQUENCE FROM N.A.
RC      STRAIN=168.
RX      MEDLINE; 97453479.
RA      Sotokun A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,
RT      Duesterhoeft A., Ehrlich S.D.;
RT      "Sequence of the Bacillus subtilis genome region in the vicinity of
RT      the lev operon reveals two new extracytoplasmic function RNA
RL      polymerase sigma factors Sigv and Sigz.";
Microbiology 143:2939-2943(1997).
[3]
CC      CC
CC      -1 SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
-----
CC      CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@lsb-sib.ch).
-----
Cc      Cc
Cc      EMBL; Y11043; CAAT1937.1; -.
Dr      DR
Dr      EMBL; U93876; AAB80898.1; -.
Dr      EMBL; Z99117; CABLA615.1; -.
Dr      HSSP; 000441; IOXA.
Cc      CC
Cc      SUBSTITUT: BG11929; CYP4.
Dr      PRAM; PF00067; P450.1.
Dr      PRINTS; PR00359; BP450.
Dr      PROSITE; PS00086; CYTOCHROME_P450; 1.
CW      Oxidoreductase; Monooxygenase; Heme-
```





DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Blotin biosynthesis; Oxidoreductase; Monooxygenase;  
 KM Electron transport; Heme.  
 FT BINDING 345 HEME (BY SIMILARITY).  
 SQ SEQUENCE 395 AA; 44865 MW; EAC3AF2637ACE1A C9C64;

Query Match	14.8%;	Score 322.5;	DB 1;	Length 395;
Best Local Similarity	27.2%;	Pred. No. 1.8e-17;		
Matches	88;	Conservative	64;	Mismatches 138;
			Indels	33;
			Gaps	7;

QY	75	YEDYRHSSECPFPRRAGCAVDIEPTSMRPEORCRALANOVGVADUKLENNRJOEL	134
Db	70	XODSHVQNMOMLEQO-----PDHRRRLATLASAGFPRRTSESOPIYIET	115
QY	135	ACSLIESLRPOGOCNFTEDYAEPRPIRIEM--LAGLPEDDIPALKYLDQMTRRPGSMT	192
Db	116	VHNLIDVOGOKKKMEVISDPA--FPLASFIANIIGVPEEDRDLKEMASLJO-----TID	170
QY	193	FAEKAELAUD-----YLPIEOROKRPTDVISIVANGOVGRPTSDAEKRMIC	242
Db	171	FTSRKALKTEGNIWAQVAMAYFKRLIOKKRRNRQOQIMSLMIGREKDK--LTDEEAASTC	229
QY	243	GLLVGLDLYVNFSLSMERLAKSPRNRDOLLERPELLRACSEBELLRRS-LVADGRIL	301
Db	230	ILAIACHETTVNLINSVYLCLODHPOLLKLRBNPDLISTAVEECIRUTSPYOMTARVA	289
QY	302	TSDFEFGVQALKKQDQILPOMLSGLDERKNACPMNHVDFROKXSHTFEGSHSLIOH	361
Db	290	SEIDIDCGVITRIGOEYUULLGANNRDPRLFTFNDRVDVDIRSPRPHLSFGHNHVCSS	349
QY	362	LARELLIVTLKEMLIRPDSIA	384
Db	350	LARLEAQIAINTLIDRMPISLNA	372

Search completed: October 4, 2000, 13:04:53  
Job time: 1684 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:39:36 ; Search time 79.83 Seconds  
(without alignments)  
359.568 Million cell updates/sec

Title: US-09-246-451-12  
Perfect score: 2179  
Sequence: 1 TTERIQSNANLAPLPHPVPE.....IVSGVOALPLVWDPATRAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL\_12:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhcc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp Vertebrate:\*  
14: sp\_undefined:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	371.5	17.0	416	2	087605 streptomyc
2	370	17.0	397	2	059523 micromonos
3	360	16.5	396	2	052544 amycolatops
4	353	16.2	410	2	09X5P9 streptomyc
5	352.5	16.2	404	2	054302 streptomyc
6	350.5	16.1	407	2	059819 streptomyc
7	332	15.2	406	2	059723 pseudomonas
8	331.5	15.2	420	2	09ZH01 streptomyc
9	327.5	15.0	376	2	031785 bacillus su
10	327	15.0	407	2	09X5P8 streptomyc
11	327	15.0	411	2	032460 actinobacteri
12	322	14.8	410	2	09X9P7 streptomyc
13	319.5	14.7	410	2	059831 streptomyc
14	318	14.6	400	2	09XDB0 mycobacteri
15	311	14.3	310	2	032927 mycobacteri
16	310	14.2	406	2	087192 streptomyc
17	309.5	14.2	397	2	024727 nocardioide
18	309	14.2	417	2	059910 streptomyc
19	307	14.1	406	2	085697 streptomyc

20	297	13.6	388	2	P77977 streptomyc
21	290.5	13.3	337	2	052816 amycolatops
22	288.5	13.2	406	2	087675 amycolatops
23	286	13.1	351	2	052572 amycolatops
24	285	13.1	421	2	052561 amycolatops
25	273.5	12.6	310	2	P96562 amycolatops
26	270	12.4	403	2	052560 amycolatops
27	268.5	12.3	411	2	09X803 streptomyc
28	267.5	12.3	366	2	059921 streptomyc
29	266	12.2	398	2	087674 amycolatops
30	265	12.2	395	2	09X5T2 streptomyc
31	258.5	11.9	437	2	09ZFC0 mycobacteri
32	255	11.7	411	2	060005 s putative
33	254.5	11.7	511	2	052569 amycolatops
34	254	11.7	406	2	052823 amycolatops
35	254	11.4	426	2	09ZFC3 mycobacteri
36	233	10.7	433	2	09XC6 streptomyc
37	231	10.6	398	2	059079 amycolata a
38	226	10.4	391	2	052822 amycolatops
39	224.5	10.3	397	2	052802 amycolatops
40	213	9.8	397	2	052802 agrobacteri
41	201.5	9.2	419	2	085655 streptomyc
42	188.5	8.7	119	2	085653 streptococ
43	185	8.5	120	2	09X418 myxococcus
44	180.5	8.3	313	2	085650 streptomyc
45	170	7.8	120	2	

## ALIGNMENTS

RESULT 1  
ID 087605 PRELIMINARY: PRT: 416 AA.  
AC 087605;  
DT 01-NOV-1998 (TREMUREL. 08, Created)  
DT 01-NOV-1998 (TREMUREL. 08, Last sequence update)  
DT 01-NOV-1999 (TREMUREL. 12, Last annotation update)  
DE CYTOCHROME P450 MONOOXYGENASE.  
GN PICK OR PIKC.  
OS Streptomyces violaceus (Streptomyces venezuelae), and  
OS Streptomyces venezuelae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC15439;  
RA BETLACH M.C., KEALEY J.T., ASHLEY G.W., MCDANIEL R.;  
RT "Characterization of the macrolide P450 hydroxylase from Streptomyces  
RT venezuelae which converts narbomycin to plicomycin.",  
RL Biochemistry 0:0-0(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S.VENEZUELAE; STRAIN-ATCC15439;  
RX MEDLINE: 99051447.  
RA XUE Y., WILSON D., ZHAO L., LIU H.-W., SHERMAN D.H.;  
RT "Hydroxylation of macrolactones XC-17 and narbomycin is mediated by  
RT the PIKC-encoded cytochrome P450 in Streptomyces venezuelae.",  
RL Chem. Biol. 5:661-669(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S.VENEZUELAE; STRAIN-ATCC15439;  
RX MEDLINE: 98445333.  
RA XUE Y., ZHAO L., LIU H.-W., SHERMAN D.H.;  
RT "A gene cluster for macrolide antibiotic biosynthesis in Streptomyces  
RT venezuelae: architecture of metabolic diversity.",  
RL Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).  
DR EMBL: AF087022; AAC64105.1; -  
DR HSBP: AF079139; AAC68886.1; -  
DR HSBP: Q00441; 10XA.  
DR PFAM: PF00067; P450; 1.  
DR PRINTS: PR00359; BP450.  
KW Monooxygenase.





```
RESULT 6
Q59819 PRELIMINARY; PRT; 407 AA.
AC Q59819;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE CYTOCHROME P450 (EC 1.14.14.1).
OLEP.
OS Streptomyces antibioticus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95255619.
RA RODRIGUEZ A.M., OLANO C., MENDEZ C., HUTCHINSON C.R., SALAS J.A.;
RT "A cytochrome P450-like gene possibly involved in oleandomycin
biosynthesis by Streptomyces antibioticus.";
RT FEMS Microbiol. Lett. 127:117-120(1995).
RL -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL; L37200; AAA9253.1; -.
DR HSSP; 000441; 10XA.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR PFAM; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT BINDING 356 356 HEME (BY SIMILARITY).
SQ SEQUENCE 407 AA; 44957 MW; 04411C60 CRC32;
```

```
Query Match 16.1%; Score 350.5; DB 2; Length 407;
Best Local Similarity 28.4%; Pred. No. 8.9e-21;
Matches 105; Conservative 63; Mismatches 167; Indels 35; Gaps 11;
```

```
QY 60 GCHHATRGQLIRAYEDYRHFSSEC---PIPREAGEANDFTPT-----SMDEPEQ 108
DB 50 GTAWLVTRMSDARIYLDSDR-FSTAATDPATPR-----MFTPEPBGVLAQDPDH 101
QY 109 ROPRALANQVGMVVDKLENRIQELACSLIESLRPOG-CNFTEDYAEPPIRIFMLLA 167
DB 102 TRLRLVKGAKFTARVDEMRPRVSLVSDLDQVANGSPADLVFLAVPRPAVICEL 161
QY 168 GLPEEDIPHLKYLTDOM---TRPDGSMTEPAKALYDYL---PIIEQRORKGTDIAIS 221
DB 162 GVPLEDDHDLRTFESDAMLSSTR---LTAAEIOHVQODFMVYMGLVAORBDAPTEDDL 217
QY 222 IVANGOVNGRPITSDKAKRMCGLLVGGLDTVNVNLSFSMEFLAKSPENHOELIERBELI 281
DB 218 ALALATNDHLLTKGELTVNMGVSLLAGHETSVNQITNLVHLLTTERKRYESLVPALV 277
QY 282 PAACEELLRRFSIYADG---RILTSDEYFPGVOLKKGQDILLPQMLSGLDERKNACPMHV 338
DB 278 PAAVEEMLRTPPLVSASFPVRATEDEVELSTVIYRAGEPCPVNHPASANRDEEVDHDEL 337
QY 339 DFRSROKVSHTTFFGSHLCLGQHLARREIIVTLKWLTRIPDSIA-PGAQIOHKSG-IV 396
DB 338 DFHRRRNPHIAFGHGACGACGAGLQELQALSAIVRRPPTLDLAPVAGLKKWQOMLI 397
QY 397 SGVOALPLVW 406
DB 398 RGLERQIVSW 407
```

RESULT 7

Q59723 PRELIMINARY; PRT; 406 AA.

AC Q59723;

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DE CYTOCHROME P450 LIN (EC 1.14.14.1).

GN LINC.

OS Pseudomonas incognita.

```
OC Bacteria; Proteobacteria.
RN [1]
RP SEQUENCE FROM N.A.
RA ROP J.D., GUNSAIUS I.C., SLIGAR S.G.;
RT Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; L23310; AAA25810.1; -.
DR HSSP; P33006; 1CPT.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR PFAM; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT BINDING 355 355 HEME (BY SIMILARITY).
SQ SEQUENCE 406 AA; 45637 MW; EFAB4E0D CRC32;
```

```
Query Match 15.2%; Score 332; DB 2; Length 406;
Best Local Similarity 26.6%; Pred. No. 2.9e-19;
Matches 106; Conservative 73; Mismatches 176; Indels 44; Gaps 10;
```

```
QY 29 YNPSNLISAGVOEAMAVIOESNVDPDLVWTRCNGHWTATRGOLREAYEDYRH-FSSRCPF 87
DB 32 WNPESDSCGF---WAVLRHNDIIE-----VSROPLFSSAYENGCHIFNENEV 77
QY 88 IPRAGEAYDFIP-TSMDPEORQOFALANQVGMVVDKLENRIQELACSLIESLRPOG 146
DB 78 GLTNAGAAVGVFFISLDPRVHTQYRKVIMPALSPARLGDIEGRIRVRALEIRIPGE 137
QY 147 QCNFTEDYAEPPIRIFMLAGL-PE-----EDIPHLKYLTDOMTRPDGS- 190
DB 138 EVDLVPLLSAPRLPLTLAELLGLDPCWYELVMTNAFVEDDEPEFKSPEDAKVYGEF 197
QY 191 MTEPAEAKALYDYLPIITFEOROKPGTDAISIYANGOVNCRPTSDPAKRMCGILLVGL 250
DB 198 MGCCQ-----ELFESRRANPGPDITFLANAEINQPALRDFIGNLTTLVGCN 247
QY 251 DTVVNLFSFMEFLANSPEHROELIERPELIPACEELLRRFSIYAD-GRILTSDEYFHG 309
DB 248 ETRRNISHTITVLSQOPDQMDILRQREPLTKTATAMVHNASPVLMRTAMEDPEIGG 307
QY 310 VOLKKGQILLPQMLSGLDERKNACPMHVDFSRQKVSHTTFFGSHLCLGHLARREIIV 369
DB 308 QALAKGDKVVLWYASGNRDSVESDADRFPVTRTGVOHVFEGSGOHVCVSRLEAMOLRV 367
QY 370 TLKEMLTRIPDFSIAPGAQIOHKSGIYGOALPLVWDP 408
DB 368 VFELISTRVAKRFELCKSR-RFRSNPLNGLKNLVVLP 405
```

RESULT 8

Q92H01 PRELIMINARY; PRT; 420 AA.

AC Q92H01;

DT 01-MAY-1999 (Tremblrel. 10, Created)

DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)

DE CYTOCHROME P450.

GN TYLH.

OS Streptomyces fradiae.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

RN [1]

RP SEQUENCE FROM N.A.

RA FOUCHES R., MELLADO E., DIEZ B., BARREDO J.L.;

RT "The left edge of the tylosin gene cluster from Streptomyces fradiae.";

RT Microbiology 0:0-0(1999).

DR EMBL; AF055922; AAD12167.1; -.

DR HSSP; 000441; 10XA.

SQ SEQUENCE 420 AA; 45531 MW; 3A3474F7 CRC32;

Query Match 15.2%; Score 331.5; DB 2; Length 420;



```

Db      46  WILTRHADVRLALADPGVSHRPGAPDQFMRNLAEI-----RAEHVLPGLLEMDPP  97
Qy      107  EDRQFALANOVGMPVVDKLENNIOELACSLIESLRQ-GOCNFTEDEVAEPPIRIEML  165
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      98  DHTKRLTLTWMFTMRARLKRLEPRTEQVETITLDMENGGSTVDLYOSFALPIPLVICE  157
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      166  LAGLEPEDIPIHLKYITDQMTPRDGSMTFAEAKA-LYDYLLPIITEORRKGTDIAISVA  224
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      158  IMGIVYEEREEFMDVLRQLQALDAPTELGAIGARMEFMYKLAAKRANCGDILLSLA  217
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      225  NGQVNGRITSDQKRMCGLLVGGDLDTVNVNLSFSMEFLAKSPENHOELIERELLPA  284
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      218  HDPPADPALTDLEIAGIGVLMILAGHETSAMKLGVGTYYTLLENDAQALLDDISLDRA  277
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      285  CEELLRRRESLVADG--RIITSDYEFHGVLQKKGOILLPQMLSGLDERRKANCPMHVDESR  342
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      278  VBEILLRHGTITVQOGLPRVYTDKMEIAGHQVYTGSSLASLPAARNRDVAVFPPDRDLITR  337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      343  QKVSHHTFGHGSHLCLGCHLARREIYTLKMLTRIRIDFDSIAPCAQ-IQHKSG-IYSGVQ  400
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      338  EHNPHLAGHGHIHLCIGELARVEMRQAMRLVTRPFGELRAAARPEDIRNRDDQIYGVY  397
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      401  ALPLVMDPA 409
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      398  NLPTVMDPA 406
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	11		
032460			
ID	032460	PRELIMINARY;	PRT; 411 AA.
AC	032460;		
DT	01-JAN-1998	(TREMBLrel. 05. Created)	
DT	01-JAN-1998	(TREMBLrel. 05. Last sequence update)	
DT	01-NOV-1999	(TREMBLrel. 12. Last annotation update)	
DE	ORF 10.		
OS	Actinomadura hibisca.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Streptosporangineae; Thermomonosporaceae;		
OC	Actinomadura.		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-PI57-2;		
RX	MEDLINE: 97480928.		
RA	DAIRI T., HAMANO Y., IGARASHI Y., FURUMAI T., OKI T.;		
RT	"Cloning and nucleotide sequence of the putative polyketide synthase		
RT	genes for pimaricin biosynthesis from Actinomadura hibisca."		
RL	Biosci. Biotechnol. Biochem. 61:1445-1453(1997).		
DR	EMBL: D87924; BAA23153.1; -.		
DR	HSSP: 000441; 10XA.		
DR	PFAM: PF00067; p450. 1.		
DR	PRINTS: PR00359; BP450.		
SO	SEQUENCE 411 AA; 44860 MW; 98A628B6 CRC32;		

Query Match	15.08;	Score 327;	DB 2;	Length 411;
Best Local Similarity	26.48;	Pred. No. 7.5e-19;		
Matches 114; Conservative	61;	Mismatches 199;	Indels 58;	Gaps 16;

[illegible]

```
QY 221 SIVANGOVNRPITSDAEKKRCGLLVGGDLTVYNELSFSEMEKLAESPHEORLEPEREL 280
      : : : : : | | | | | : : : : :
Db 222 SILSRHADD-GLSEEBELLPGMTLLAAGYHTTTAAITHSVHLRREPSRYARLRDEPSC 280
      : : : : : | | | | | : : : : :
QY 281 IPAAEELLRRPSLVAOS---RLTSDYERHQVLAKKGQIILLPOLSLDERKNACPM 336
      ||||| ||||| : : : | | | : : : : :
Db 281 IPAAAEELL-RYGOGGGGAGAIRAVEDEVGGTLVRAGEAVIPLEFNAARNDVEVPADPE 339
      ||||| ||||| : : : | | | : : : : :
QY 337 HVDESROKVSHTTFGCHSGHLCOHLARRELIYVLKEMLTRIDPFSA-PGAOIHKSGI 395
      : : : : : | | | | | : : : : :
Db 340 EDDGRDGNPHIALGHGHGYCLGARPLARELOVVLELTVERTALRLAIADADITMRPGL 399
      : : : : : | | | | | : : : : :
QY 396 V-SGVQALPLVM 406
      : : : : : | | | | | : : : : :
Db 400 AFARADALPIAM 411
```

RESULT	12			
Q9X9P7				
ID	Q9X9P7	PRELIMINARY;	PRT;	410 AA.
AC	Q9X9P7;			
DI	01-NOV-1999	(TREMBLrel. 12, Created)		
DI	01-NOV-1999	(TREMBLrel. 12, Last sequence update)		
DI	01-NOV-1999	(TREMBLrel. 12, Last annotation update)		
DE	NIFE PROTEIN.			
GN	NIFE.			
OS	Streptomyces tendae.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-TUE901;			
RA	BRUNTNER C., LAUER B., SCHWARZ W., MOEHRL V., BORMANN C.;			
RT	"Molecular characterization of co-transcribed genes from Streptomyces			
RT	tendae T901 involved in the biosynthesis of the peptidyl moiety of			
RT	the peptidyl nucleoside antibiotic nikkomycin.";			
RL	Mol. Gen. Genet. 0:0-0(0).			
SR	EMBL; Y18574; CAB46536.1; -			
SQ	SEQUENCE	410 AA;	45084 MW;	9B52CB74 CRC32;

Query Match	14.8%	Score 322;	DB 2;	Length 410;
Best Local Similarity	27.3%	Pred. No. 1.9e-18;		
Matches 99; Conservative	64;	Mismatches 179;	Indels 20;	Gaps 9;

[illegible]

RESULT	13		
059831			
ID	059831	PRELIMINARY:	PRT: 410 AA.
AC	059831;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)		
DE	CYCROCHROME P450 SCA-2 (EC 1.14.14.1).		
OS	Streptomyces cardophilus.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;		
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-SANK 62585;		
RC	MEDLINE; 96001248.		
RA	WATANABE I., NARA F., SERIZAWA N.;		
RT	"Cloning, characterization and expression of the gene encoding		
RT	cytochrome P-450sca-2 from Streptomyces cardophilus involved in		
RT	production of pravastatin, a specific HMG-coA reductase inhibitor."		
RL	Gene 163:81-85(1995).		
DR	EMBL; D30815; BAA06492.1; -.		
DR	HSSP; Q00441; 10XA.		
DR	PFAM; PF00067; P450. 1.		
DR	PRINTS; PR00359; BP450.		
SO	SEQUENCE 410 AA; 45049 MW; 2DF70C8E CRC32;		

Query Match	14.78;	Score 319.5;	DB 2;	Length 410;
-------------	--------	--------------	-------	-------------

Best Local Similarity 25.9%; Pico. NO. 3e-16;  
Matches 93; Conservative 63; Mismatches 184; Indels 19; Gaps 7;

QY	63	WIAHQGLIRAYADYR-----HNSSEKPLP--REAGAEADLPRTSMDEPQOFR	112
Db	56	WVTKHEAARLADPRLLSSDRCLHADPRAPSREKARQSPAF---IOMDPREHTR	111
QY	113	ALANQVGMRYVDLLENRIQELASLLESRLPQS-QCNFEDNAERPRPIPIFILLAGRE	177
Db	112	RMTISEFTVKRIKQMRPDERIVTHGFIIDMLIAGRTYDALVSQELRPVSNVICHMLGVPY	177
QY	172	EDPHLKYLTDOMTRPRGSMFAFEAKELDYLLRPIAREORAKPGTDAISIVANGUYNR	231
Db	172	ADHEFPDASKRLQVAWDADSAVAARDFEORLYDGLITKLESSEGTGLSKLYTHQADG	231
QY	232	PITDEAKRMCGLLLVGGDLTVVNFSLFSMEGLAKSPENHOELERELLIPACEELLRR	291
Db	232	EIDRAELISTALLLVVNGHETTASMTLSVYTTLLENDDNALAEADSLVGAEEELLRV	291
QY	292	FSL--VADGRILSDYEHGQVLKKQDQILPOMLSGLDERKNACRYNHPFESROKVSHT	349
Db	292	LAIADIAGRIATADDEIDQQLIAGBGVLYTNSIARSDSVFENPRDRLVHNSARNHLS	351
QY	350	FGHSHSLCSGHLARRELYTLAKSMILTRIPFSLA-FGAOLQKSG-IYSGVADALRYW	406
Db	352	FGYHGOHLCGNLARLELYTLVLEDRIPRLRYLAUVEODITLTPRGTTIGGVNELRTW	410

RESULT	14
09XDB0	
ID	09XDB0
AC	09XDB0;
DT	01-NOV-1999 (TREMBLrel. 12, Created)
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE	CYTCHROME P450.
GN	P450.
OS	Mycobacterium smegmatis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-MC2155;

RX MEDLINE: 99287823.  
 RA POUPIN P., DUCROCQ V., HALLIER-SOULIER S., TRUFAUT N.,  
 RT "Cloning and characterization of the genes encoding a cytochrome P450  
 RT (PipA) involved in piperidine and pyridoline utilisation and its  
 RT regulatory protein (PIPR) in *Mycobacterium smegmatis* mc2155.".  
 RL J. Bacteriol. 181:3419-3426(1999).  
 DR EMBL: AF102510; AAD28344.1;  
 QO SEQUENCE 400 AA: 44747 MW: C10D01A CRC32:

Query match	Score	DB 2;	Length
14.68;	318;	DB 2;	400;
27.08;	318;	DB 2;	400;
27.08;	318;	DB 2;	400;

Matches 116; Conservative 57; Mismatches 170; Indels 86; Gaps 14;  
 best local similarity 27.06; Pired. NO. 3.9E-18;

QY	26	EDMNPN--SNLSACVOEA--W-----AVLOESVNDLV-----WTRCNG	60
Db	12	FDVYIDPASTISDEVHEHREKSWATTPYGLAVLRQYQVNRLLIHNRLRGOSAMPAHNG	71
QY	61	-----GHMIATRGQILREAVEDYRHFSSECPFIIPREAGEAYDEIPTSMDDPEOROFAL	114
Db	72	VTEGPEAFEMFAS-----WILNKEGE-----EHNRRL	99
QY	115	ANOVYGMVVDKLENNIOELACSLIESLRPOGOCNPFEDAEPPRIITPMILAGLPREDI	174
Db	100	MNPAPSPKLISLVPRCALANELLIDNFAEPDREFEYSEAEPPAAVIAIMGLPEE-	158
QY	175	PLKYLITDQ-----MTRPDSMFEAKELADYULPIIQROROKGTDAISIVAN	225
Db	159	--WKVISTESATIGALGVTLREDLPKEAENVQRLYEXSDELLIDNRANRRDPMITLVN	216
QY	226	GOVNGRP-----ITSDEAKRMCGLLVGGDLTVNPFLSMEFLKSPDHNOELIERPELI	281
Db	217	A---SRPDDGRLSKEURDATALILILFGGFDTTRQOLGLAMOTFEKHNPDQMRLLGEREDLG	273
QY	282	PACCELLRRFSLSVA--DGRILTSYEFHGYOLKKGDOILLPQMSGDERKNACPMHVD	340
Db	274	GRAYEEVNRKVPYRWYTRREVEDFEYEGVYLKKGTYVHLYSESAGTDPR--VFEPGFDI	333
QY	341	SRQAVSHTFEGHSHLCIGOHNAAREIITVLEKMLTRIPESTIAPGAQIOHKSGIVSGVO	400
Db	332	TAERKPRHFGFCGCVHNCIGHFVARSDSMEALPRLLARLRDPEHNELRGATWLPDSC--NTGPN	390
QY	401	ALPLVMDPA 409	
Db	391	TLPIGFTPA 399	

RESULT 15  
032927

ID	032927:	PRELIMINARY:	PRT:	310 AA.
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-NOV-1998 (TREMBLrel. 12, Last annotation update)			
DE	CYTCHROME P450 (FRAGMENT).			
GN	MLCB1788.51C.			
OS	Mycobacterium leprae.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;			
OC	Actinomycetales; Corynebacteriinae; Mycobacteriaceae; Mycobacterium			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	SKELTON J., CHURCHER C.M.;			
RL	Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	PARKHILL J., BARRELL B.G., RAUANDREAM M.A.;			
RL	Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 93188700.			
RT	EIGLMEITER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;			
RT	"Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae.";			





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:35:16 ; Search time 76.18 Seconds  
(Without alignments)  
128.722 Million cell updates/sec

Title: US-09-246-451-13

Perfect score: 2177

Sequence: 1 TTETIQSNANLAPLPHPHVE.....IVSGVQALPLVWDPATRKAV 414

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues  
Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2122	97.5	405	1	Y04128	Pseudomonas putida
2	1185	54.4	446	1	Y04126	Bacterial and mamm
3	362	16.6	411	1	R60777	Mycilamycin IV hyd
4	337	15.5	411	1	W54389	Actinomadura hibis
5	328.5	15.1	587	1	W33274	S. fradiae tylosin
6	328	15.1	406	1	R11349	Cytochrome enzyme
7	324.5	14.9	398	1	W11585	Streptomycetes prist
8	322	14.8	403	1	R11350	Cytochrome enzyme
9	321	14.7	408	1	R77867	S. clavuligerus OR
10	314	14.4	412	1	R38309	Sequence of the p4
11	311.5	14.3	410	1	R51368	Protein containing
12	294.5	13.5	404	1	R14724	6-hydroxylase enco
13	190	8.7	398	1	R47521	Vitamin D hydroxyl
14	161	7.4	422	1	W36128	Daunomycin C-14 hy
15	161	7.4	474	1	W36132	N-terminal modifie
16	156	7.2	422	1	M00729	Daunorubicin 14-hy
17	152	7.0	533	1	R15057	Cytochrome P450C25
18	142	6.5	494	1	R62825	Human steroid-21-h
19	135.5	6.2	1169	1	R76544	Mitochondrial cyto
20	133	6.1	518	1	W67616	A. nidulans phenyl
21	131	6.0	512	1	W93216	Human cytochrome p
22	129	5.9	508	1	W35711	Chrysanthemum flavon
23	129	5.9	512	1	R72365	Human auxillary cy
24	128	5.9	512	1	R93172	Human cytochrome p
25	128	5.9	512	1	R72366	Human auxillary cy
26	128	5.9	512	1	W00652	Cytochrome P4501A1
27	128	5.9	512	1	R93173	Human cytochrome p
28	125	5.7	524	1	P70577	Rat hepto-cytochr
29	123	5.6	503	1	R72363	Human cytochrome p
30	123	5.6	503	1	R81464	Human derived cyto
31	123	5.6	503	1	R93170	Human cytochrome p
32	123	5.6	503	1	Y05202	Human CYP3A4 prote
33	123	5.6	512	1	R72364	Human auxillary cy

34	123	5.6	512	1	R93171	Human cytochrome p
35	122.5	5.6	523	1	R59291	Rat liver cytochro
36	122.5	5.6	898	1	P61030	Entire coded sequ
37	122.5	5.6	898	1	P61082	Complete translati
38	122.5	5.6	899	1	P61056	Translation of pla
39	122.5	5.6	1144	1	P81334	Expression prod. o
40	122.5	5.6	1150	1	P81335	Expression prod. o
41	122.5	5.6	1150	1	P81337	Expression prod. o
42	122.5	5.6	1162	1	P81336	Expression prod. o
43	121	5.6	514	1	W34539	Cytochrome P450 cc
44	120.5	5.5	493	1	R72362	Human cytochrome p
45	120.5	5.5	493	1	R81467	Human derived cyto

ALIGNMENTS

last-processed

RESULT 1	Y04128 standard; Protein; 405 AA.
ID Y04128;	
AC 11-JUN-1999 (first entry)	
DE Pseudomonas putida cytochrome P450 protein P450-cam.	
KW Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;	
KM oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;	
KW bioremediation; environmental pollutant.	
OS Pseudomonas putida.	
PN M09908812-AL.	
PD 25-FEB-1999.	
PF 17-AUG-1998; 016979.	
PR 20-AUG-1997; US-056754.	
PA (UYRP ) UNIV ROCHESTER.	
PI Jones JP, Shimoji M;	
DR WPI: 99-190131/16.	
DR N-PSDB: X19926.	
PT New P450 fusion proteins - comprising a portion of a bacterial	
PT cytochrome P450 protein and a portion of a mammalian cytochrome P450	
PT protein	
PS Disclosure: Page 12-13: 51pp; English.	
CC The present invention describes a fusion proteins comprising a portion	
CC of a bacterial cytochrome P450 protein and also a portion of a mammalian	
CC cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or	
CC any compound having a carbon-hydrogen bond. The fusion protein can be	
CC used for hydroxylating a compound to be oxidised. It can also be used in	
CC the bioremediation of an environmental pollutant. Since the fusion	
CC protein is soluble, it can be subject to structural elucidation by X-ray	
CC crystallography for designing functional proteins. It can be readily	
CC expressed in soil bacteria to facilitate bioremediation. The present	
CC sequence represents Pseudomonas putida cytochrome P450 protein P450-cam	
CC from the present invention.	
SO Sequence 405 AA:	
Query Match	97.5%; Score 2122; DB 1; Length 405;
Best Local Similarity	99.5%; Pred. No. 2e-210;
Matches 403; Conservative 1; Mismatches 1; Indels 0; Gaps 0;	
Db 10 NLAPLPHPHVEHLVDFDPMTPSNLSAGVDEAMAVLDGSNPDLVWTRCNGHMIATRQ 69	
Db 1 NLAPLPHPHVEHLVDFDPMTPSNLSAGVDEAMAVLDGSNPDLVWTRCNGHMIATRQ 60	
QY 70 LIRAYEDYRHFSSECFPIPEAGEAYDFPTSMDDPEOROFRLANOVGMPVVDLEN 129	
Db 61 LIRAYEDYRHFSSECFPIPEAGEAYDFPTSMDDPEOROFRLANOVGMPVVDLEN 120	
QY 130 RIQELASLIESLRPOGQCNTEDEYAEFPPIRIMLAGLPEEDIPHLKYLTDQMPD 189	
Db 121 RIQELASLIESLRPOGQCNTEDEYAEFPPIRIMLAGLPEEDIPHLKYLTDQMPD 180	
QY 190 SMTPAEAKKALYDLIPITIDOROKPGTDAISIVANGVNGRPITSDAKRMGCLLVG 249	
Db 181 SMTPAEAKKALYDLIPITIDOROKPGTDAISIVANGVNGRPITSDAKRMGCLLVG 240	

OY	250	LDTVNLFSESNBEFLAKSPSEHROELIERPERIPACEELLRRSLVADGRILTSQVEFHG	309
DB	241	LDTVNLFSESNBEFLAKSPSEHROELIERPERIPACEELLRRSLVADGRILTSQVEFHG	300
OY	310	VOLKKGDOILLPQMLSGLDERKNACPMHVDFSQKVSHTTFCGSHLCGLQHAREIIV	369
DB	301	VOLKKGDOILLPQMLSGLDERKNACPMHVDFSQKVSHTTFCGSHLCGLQHAREIIV	360
OY	370	TLKEWLTIRIPDESIAPGAQIOHKSGIVSQVALPLVWDATPTKAV	414
DB	361	TLKEWLTIRIPDESIAPGAQIOHKSGIVSQVALPLVWDATPTKAV	405
RESULT	2		
ID	Y04126		
AC	Y04126	standard; Protein; 446 AA.	
DT	11-JUN-1999	(first entry)	
DE	Bacterial and mammalian chimeric cytochrome P450 protein.		
KW	Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;		
KW	oxidase; hydrocarbon; carbon-hydrogen bond; hydroxylating;		
OS	bioremediation; environmental pollutant.		
OS	Synthetic.		
PN	WO9908612-A1.		
PD	25-FEB-1999.		
PF	17-AUG-1998; U16979.		
PR	20-AUG-1997; US-056754.		
PA	(UYRP ) UNIV ROCHESTER.		
PI	Jones JP, Shimoji M;		
DR	WPI: 99-190131/16.		
DR	N-PSDB: X19916.		
PT	New P450 fusion proteins - comprising a portion of a bacterial		
PT	cytochrome P450 protein and a portion of a mammalian cytochrome P450		
PS	protein		
PS	Claim 24; page 6-8; 51pp; English.		
CC	The present sequence is a fusion proteins comprising a portion of a		
CC	bacterial cytochrome P450 protein and also a portion of a mammalian		
CC	cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or		
CC	any compound having a carbon-hydrogen bond. The fusion protein can be		
CC	used for hydroxylating a compound to be oxidised. It can also be used in		
CC	the bioremediation of an environmental pollutant. Since the fusion		
CC	protein is soluble, it can be subject to structural elucidation by X-ray		
CC	crystallography for designing functional proteins. It can be readily		
CC	expressed in soil bacteria to facilitate bioremediation.		
Sequence	446 AA;		
30			

Query Match	54.4%	Score 1185	DB 1	Length 446
Best Local Similarity	58.0%	Pred. No. 9e-114		
Matches 257	Conservative 32	Mismatches 92	Indels 62	Gaps 11
Qy 10	NLAAPRPHVPEHLVDFDMVNPNSLSAGQOEAWALQESNVVDLWTRNGGHWIATRGO	69		
Db 1	NLAAPRPHVPEHLVDFEDMTNPNSLSAGQOEAWALQESNVVDLWTRNGGHWIATRGO	60		
Qy 70	LIREAVEDYRIHSSSECPFIPREAGEAYDFIPASMPDPEOROPALANOVGAPVYDKLEN	129		
Db 61	LIREAVEDYRIHSSSECPFIPREAGEAYDFIPASMPDPEOROPALANOVGAPVYDKLEN	120		
Qy 130	RIQELACSLIESLRQGGCNFTEDYAEERPIKIFMLLGLPREDIPIHLKYLLIDOMKTRPG	189		
Db 121	RIQELACSLIESLRQGGCNFTEDYAEERPIKIFMLLGLPREDIPIHLKYLLIDOMKTRPG	180		
Qy 190	SMTFAEAKEAALDYLIPIIEQRQRPGT-----DAISIVANGQVNGRP-----ITSDE--	237		
Db 181	SMTFAEAKEAALDYLIPIIEQRQRPGT-----DAISIVANGQVNGRP-----ITSDE--	240		
Qy 238	AKRMFGILLVGGDLTVNVFLSFSMEFLAKSPREHQELIERPERI-----	281		
Db 241	AVLDLFG---AGTEFTTSTLRYALLLLKHEEVPAKQGEIERIYVGRNRSPCMQRSHMP	296		
Qy 282	--PACCELLRRFRSLVADG--RIILTSDEFNIGVOLKKGQIIL-----PQM 323			

```

Db      297 YTDVAHVNEVQKIDLLFTSLPHAVTCDIKRANLIRKGTITILSTLSVLHDKKEPNEM 356
Oy      324 LSG-----LDERRKNACPMHWDEFSROKVSHHTFGHGSHLCGLQHARREIIVTLKEWLTIPD 380
          ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      357 FDPHHFLFDEGNN-----FKSKY-FMPSACKRICVEALAGMLFLLTSILQNPL 408
Oy      381 FSIAPGAQLQHKSKIYSVGQALP 403
          | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      409 KSLVDPKNLD-TTPDVNVNGFASVP 430

RESULT      3
R60777
ID      R60777 standard; Protein; 396 AA.
AC      R60777;
DT      21-JUN-1995 (first entry)
DE      Mycinamicin IV hydroxylating protein.
KW      Mycinamicin; hydroxylase; macrolide; antibiotic; Micromonospora;
KM      Micromonospora griseorubida.
OS      Micromonospora griseorubida AT11725CN3.
PN      J06253853-A.
PD      13-SEP-1994.
PE      09-MAR-1993; JP-047638.
PR      09-MAR-1993; JP-047638.
PA      (ASAHI ) ASAHU KASEI KOGYO KK.
DR      WPI: 94-328997/41.
NR      N-PSDB: 073674.
PT      DNA encoding a protein having mycinamicin IV hydroxylating
        activity - for prodn. of mycinamicin, a macrolide antibiotic
ES      Claim 1; Page 12-14; 23pp; Japanese.
CC      The amino acid sequence of a protein having mycinamicin IV hydroxylase
CC      (MH) activity. The gene encodes a protein of 396 a.a. The DNA was
CC      obtained from the macrolide antibiotic-producing bacterium Micromonospora
CC      griseorubida A11725CN3/pRYS507. The gene was isolated from the plasmid
CC      pRYS507. The protein encoded by this plasmid can be used to produce
CC      mycinamicin IV in pRYS507-deficient Micromonospora strains.
SO      Sequence 396 AA;

```

Query Match	16.6%	Score 362	DB 1	Length 396
Best Local Similarity	28.5%	Pred. No. 5, 2e-29		
Matches 105	Conservative	57	Mismatches 166	Indels 38
Gaps				8
QY	63	WIATQGILIREAYEDYRIHFSSECPFI-----PRAGEAYADFIPISMDPPEQKQFRA	113	
DB	43	WLVTR-----YEDVRAVLGDGRVGRGSPMTRDERKTRPEAWKGLSMDPEHSRLRR	95	
QY	114	IANQVGNPVPYDKLENRIQELACSLIESLRPOGQ-CNFTEDYAEPPPIRIFMLAGLPEE	172	
DB	96	LIVKAKFTARRAESLPRPRAEIHAEIIVDQMAAGCPADLVAMPARQLPVAVICELLGVP5A	155	
QY	173	DIPHLKYLTDQMTDRDGSN-----TFAAKAEALYYLPIIPIEOROKQGTDAISIV	223	
DB	156	D-----HRTFRKSGAVLSTAEVTAEEOMNABQAVAYYMDLIDRRKKEPTDDLVSAL	208	
QY	224	ANGQVNGRPITDEDEAKRMFGLLVGGLDVPVNFLSFSMEFLAKSPHROELIERPERIA	283	
DB	209	VQARQOQBSLSQOELLDLAIGLLVAGYESTTQIADFYVLLMTREPLRKQLLDRPLLTS	268	
QY	284	ACEELLRRFSL--VADGRILTSDEYFHGVSOLKKQDQILLPQMLSGDLERKNACPMWYDF	340	
DB	269	AVEELTRWVPLGVGTAFPRYAVEDYTLRGVTLRAGEPYLASSGANNRDQAQPPDADRIDV	328	
QY	341	SRQAKSHHTFGGSHLCIGOHARREIYITLKEMLTIRIPDSISA-PGAOIQKSG-IVSG	398	
DB	329	DTPNQHGFGHGIVHCHGAPLARVELQVALFVLLQRLRPGILGIPETQIRSESGMLTGG	388	
QY	399	VQALPLVW 406		
DB	389	PLELPVW 396		
RESULT	4			





PS Claim 15; page 158; 224pp; English.  
CC This cytochrome P450 enzyme, P450S2, is expressed alongside the iron  
CC sulphur protein P450-A, by a DNA sequence contained in a recombinant  
CC plasmid. Host Streptomyces species are transformed with the plasmid  
CC and are used to coat a plant seed to transform the plant. The res-  
CC ultant transformants are resistant to herbicides.  
CC See also Q1126.  
SQ Sequence 403 AA;

Query Match 14.8%; Score 322; DB 1; Length 403;  
Best Local Similarity 28.4%; Pred. No. 7; 1e-25;  
Matches 103; Conservative 64; Mismatches 166; Indels 30; Gaps 13;

QY 63 WIRPGQLIEAEVDYHFSSEC-----PIIPREAGAYDFIPI--SMDEPQROFALA 115  
DB 52 WLVTRHODVAVIGDPR-ESADAHRCGFPPLTMGGRITIGTNPFLMDDPEHARLRML 110  
QY 116 NOVGMFVVDKLENRIQELACSLIESLRP-OGOCNFTEDYAEPPPIRIFMLAGLPEEDI 174  
DB 111 TADFIYKKEVAMREYQRLADLDVDRMTGRTSADLVTERALPLSLVICLLGVPEDH 170  
QY 175 PHL-----KYLTDMTRPDGSMTFEAKEALYDYLPIIEOROKPGTDAIS-IVANGOVN 229  
DB 171 AFROERSRVLTLSRSTPE---EVRADDELLEYLAVLARLKRERPDALISRLVARGELD 227  
QY 230 GRPITSDSEAKRMFGILLVGLDVFVNFLSFMELAKSPERHQLIERPERIRPAACEEL 289  
DB 228 DTQIAT-----MGRLLVAVARGELDDQIATMGLLRNPOLARLRAEPALVGAVELL 282  
QY 290 RRESLVADG--RILTSDEYFHVOLKKGDQILLPOMLSGLDERKNACP--MHVDFSRQV 345  
DB 283 RYLTIYVNGVPRIATEVEVLGGRTIAGEGVLC--MISSANRAEVPFGDDLDVADAR 340  
QY 346 SHTFPGHSLICGOLARREIIVTKEMLTRIPDSIA-PGAQIQKSGI-VSGVALP 403  
DB 341 RHVAFGGVHQCIGQPLARVELQIAIETLLRLRLRLAVPHEEIPFGDMATGVHSLP 400  
QY 404 LVW 406  
DB 401 IAW 403

RESULT 9  
R77867  
ID R77867 standard; Protein; 408 AA.  
AC R77867;  
DT 13-NOV-1995 (first entry)  
DE S. clavuligerus ORF10 product.  
KM Clavulinate acid; clavulinate; antibiotic; beta-lactamase-inhibitor.  
OS Streptomyces clavuligerus.  
PN CA2108113-A.  
PD 09-APR-1995.  
PF 08-OCT-1993; 108113.  
PR 08-OCT-1993; CA-108113.  
PA (UYAL-) UNIV ALBERTA.  
PI Aidoo KA, Jensen SE, Paradkar AS;  
DR WPI: 95-207301/28.  
DR N-PSDB: Q91580.  
PT Clavulinate acid biosynthesis enzymes and corresp. DNA - useful for  
PT biosynthesis of the antibiotic in Streptomyces hosts which do not  
PT naturally produce clavulinate  
PS Claim 32; Fig.19; 41pp; English.  
CC A 15 kb fragment S. clavuligerus NRRL 5741 genomic DNA (Q91580),  
CC extending downstream from pccC, included 10 ORFs encoding the  
CC enzymes required for clavulinate biosynthesis. The ORF10  
CC product (R77867) showed high similarity to cytochrome P450-type  
CC enzymes from other Streptomyces spp.  
SQ Sequence 408 AA;

Query Match 14.7%; Score 321; DB 1; Length 408;  
Best Local Similarity 28.8%; Pred. No. 9; 2e-25;

Matches 121; Conservative 59; Mismatches 186; Indels 54; Gaps 18;

QY 16 PHEPHLVPDFDMWNPENSLGVOEAMAVIOESVPLVWTRCGGH-WIATROQLIREA 74  
DB 14 PAVPMHVCYVD---PPQLAGLSOKAASRYT-----LW---DGSQVWLTSAGARAV 62  
QY 75 YEDYRHS-SECFIP-----REAGYDFIPTSMDEPQROFRA-----LANOV 118  
DB 63 LGDRFTAVYSAPGFPMLTRTSQLVANPESASTI--RMDDPQSRLSRSLTRFLARRA 120  
QY 119 VGM-PVVDKLENRIQELACSLIESLRPQOCNFTEDYAEPPPIRIFMLAGLPEEDI 177  
DB 121 EALRPVAREL---IDELIGLVKGERP---VDLVAGLTIPPSRVITLLGAGCDRREFI 174  
QY 178 K-----YLTDMTRPDGSMTFEAKEALYDYLPIIEOROKPGTDAISIVANGOVNRP 233  
DB 175 EDRSAVLIDGTYPE---QVAKARDELIDGYLREIVEERIEPGLDLSRLVIDQVRGHL 231  
QY 234 TSDEAKRMFGILLVGLDVFVNFLSFMELAKSPERHQLIERPERIRPAACEELRFS 293  
DB 232 RVEEMVPMKULLVAGHGTTSQASLSLSLNDPELAGRUTEDRPALLPRAVEELRFS 291  
QY 294 LVADG--RILTSDEYFHVOLKKGDQILLPOMLSGLDERKNACPMDVDFSRQVSHTFG 351  
DB 292 IVQGLARAAVEDVDLDVILRAGEGVYLSLGNRDETVLPDRVDVDRDARRHLAFG 351  
QY 352 HGSMLICGOLHAR---REIIVTKEMLTRIPDSIA-PGAQIQKSGI-VGVALPVW 406  
DB 352 HGMHQLGOMLARVELEILLAAVLRW---PGARLAVPEELDRHREVSIGCALPVTW 408

RESULT 10  
R38309  
ID R38309 standard; Protein; 412 AA.  
AC R38309;  
DT 04-DEC-1993 (first entry)  
DE Sequence of the P450-soy protein.  
KM P450soy; soy; gene; oxidation; haem protein.  
OS Streptomyces griseus ATCC 13273.  
PN W09312236-A.  
PD 24-JUN-1993.  
PF 16-DEC-1992; U10885.  
PR 16-DEC-1991; US-807001.  
PA (DUPO ) DU PONT DE NEMOURS & CO E.I.  
PI Omer CA, Sariaslani FS, Trower MK;  
DR WPI: 93-214178/26.  
DR N-PSDB: Q45569.  
PT Constitutive expression of P450 SOY (SOYC) and ferredoxin soy  
PT (soyD) in Streptomyces - used for oxidn. of organic chemicals  
PS Example: Figure 2; 45pp; English.  
CC Cytochrome P450soy was purified from S.griseus ATCC 13273. Two  
CC similar forms of P450soy were isolated. P450soy-delta, is derived  
CC from P450soy by in vitro proteolysis during isolation. One of the  
CC tryptic peptide fragments of cytochrome P450soy and of of the  
CC P450soy-delta protein were subjected to automated degradation. The  
CC NH2 terminal sequences are given in R38306 and R38307. A mixture of  
CC oligos that consist of possible DNA sequences that could encode the  
CC AAs GGVHQC of the tryptic peptide was made. It consists of the  
CC sequences in Q43290-23. The oligo mixture was end-labeled and used  
CC to probe the EMBL4 library of S.griseus DNA. Hybridising plaques  
CC were isolated and a 4.8kb SacI DNA fragment was isolated from one  
CC clone that hybridised to the oligo probe mixture. As segment of the  
CC 4.8kb fragment was sequenced and found to contain an ORF. Within  
CC this ORF was a section that corresp. exactly to the AA sequence  
CC determined from the cytochrome P450soy tryptic peptide (see Q45569,  
CC R38309). The gene encoding the P450soy protein was called soyC. Five  
CC nucleotides downstream from the stop codon for soyC another ORF was  
CC identified. This ORF encodes an apparent ferredoxin-like protein.  
CC The gene was designated soyB and the protein ferredoxin-soy.  
SQ Sequence 412 AA;

Query Match 14.4%; Score 314; DB 1; Length 412;





KW Daunomycin C-14 hydroxylase; doxa gene; doxorubicin; daunomycin;  
KW 13-dihydrodoxorubicin; daunomycin; anthracycline;  
KW anticancer; cytostatic; cancer; therapy; plasmid pANT199.  
OS Chimeric - streptomycetes sp. strain C5.  
OS Chimeric - synthetic.  
FH Key Location/Qualifiers  
FT Cleavage-site 31  
FT Protein /note="enterokinase cleavage site"  
FT 53..474  
PN W09744439-A2. /note="native daunomycin C-14 hydroxylase"  
PD 27-NOV-1997.  
PE 22-MAY-1997; 008690.  
PR 24-MAY-1996; US-653650.  
PA (OHIS ) UNIV OHIO STATE RES FOUND.  
PI Desanti CL, Dickens ML, Strohl WA;  
DR WPI: 98-018495/02.  
DR N-PSDB: V01452.  
PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase  
PT - also hydroxylation and oxidation of other anthracycline(s) with  
PT the same enzyme  
PS Disclosure: Fig 9; 59pp; English.  
CC This protein comprises an N-terminal modified enzyme derived from  
CC the daunomycin C-14 hydroxylase (see W36128) of Streptomyces sp.  
CC strain C5. It is encoded by a gene construct (see V01452) in  
CC plasmid pANT199. In this plasmid, the doxa gene (see V01447) is  
CC translationally fused with a leader sequence encoding 6  
CC histidine residues so that the fusion protein can be affinity  
CC purified on a nickel-agarose gel. Daunomycin C-14 hydroxylase  
CC is a P450-like enzyme capable of converting daunomycin to the  
CC anticancer agent doxorubicin. Host cells, especially Streptomyces  
CC host cells, transformed with pANT199 can be used in methods for  
CC the production of doxorubicin from daunomycin or for the  
CC hydroxylation and oxidation of other anthracyclines.  
SO Sequence 474 AA;

Query Match 7.4%; Score 161; DB 1; Length 474;  
Best Local Similarity 23.6%; Pred. No. 3.6e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 QEAMAVIQESNVDPVTRNGGHIMATRGQILREATEDYRHSSF--CPFIPEAGEAY 96  
DB 83 REAGPVV-EVNAP-----AGGPAMVITDDALAREVLADRFVKGPDLPATAMRGVDDGL 135  
QY 97 D-----FIPSMPPPEOROPALANQVGMVVDKLENRIQELACSLIESL-----R 143  
DB 136 DIVPELRPPTLAVDGEDHRLRRIAPAFNFRRLAERTDRIATADRLITELADSSDR 195  
QY 144 PGGOCNFTEDYAEFPPIRIFMLAGLP-----EEDIPHLKYL-----TDQMTRPD 188  
DB 196 SGPEAELIGGFAYHFLVLCGLGVPTDPAMAREAVGVKALGLGQPQAGDGTDP 255  
QY 189 GSKTFEAKFALDYLPITTEORQRKGTDAISIVANGQVNGRPTSDEAKRMFGLLVG 248  
DB 256 GDVPDTSALTESL--LEAVAHARRKDRITRVLYERAQAFEGSVSDQLVYMITGLIFA 313  
QY 249 GLDPTVNFLSFMSFELAKSPENROELIERPERIPACCELLRR-----FSIVADGRILTS 303  
DB 314 GHDTTGSFLGF---LLAEVLAGRLADADGDAISRFEALRHHPVPSL---WRFAAT 367  
QY 304 DYFHHGVQLKKDQILLPQMLSGLDERKNACPMHVDSPKQVSHHTTGGSHLCLGQHILA 363  
DB 368 EVVIRGVRLPRGAPVLDIEGTMTDGRHNHDAPAHFHPDRSRRRLTFPGDPRHYCIGBOLA 427  
QY 364 RRELIYLIKEMLIRIPDFSLA--PGAQIQ--HKSGIVSGVALPLVW 406  
DB 428 QLESRTMIGVLRSRFPQARLAVPYEELRMCRCGAQTARLTDLR-VW 472





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:38:07 ; Search time 66.92 Seconds  
(Without alignments)  
103.693 Million cell updates/sec

Title: US-09-246-451-13  
Perfect score: 2177  
Sequence: 1 TTERIQSNANLAPLPHPVE.....IVSGVQALPLVWDPATTKAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 164575 seqs, 16761186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgnl\_7/prodata/1/iaa/5A.COMB.pep:\*  
2: /cgnl\_7/prodata/1/iaa/5B.COMB.pep:\*  
3: /cgnl\_7/prodata/1/iaa/5.COMB.pep:\*  
4: /cgnl\_7/prodata/1/iaa/PCVUS.COMB.pep:\*  
5: /cgnl\_7/prodata/1/iaa/backfillsl.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	363.5	16.7	416	US-09-320-878-18	Sequence 18, Appl
2	340	15.6	403	5212296-9	Patent No. 5212296
3	328	15.1	406	5212296-6	Patent No. 5212296
4	314	14.4	412	US-08-102-863-11	Sequence 11, Appl
5	314	14.4	412	PCT-US92-10885-11	Sequence 11, Appl
6	271.5	12.5	419	US-09-335-409-8	Sequence 8, Appl
7	161	7.4	422	US-09-096-982-5	Sequence 5, Appl
8	161	7.4	422	US-08-653-650A-5	Sequence 5, Appl
9	161	7.4	422	US-09-096-982-8	Sequence 8, Appl
10	158	7.3	443	US-09-096-982-9	Sequence 9, Appl
11	158	7.3	443	US-08-653-650A-9	Sequence 9, Appl
12	158	7.3	443	US-08-396-218-2	Sequence 2, Appl
13	156	7.2	422	US-08-760-116-2	Sequence 2, Appl
14	156	7.2	422	US-08-194-981E-5	Sequence 5, Appl
15	131	6.0	512	US-09-320-878-7	Sequence 7, Appl
16	123.5	5.7	490	US-08-201-118-3	Sequence 3, Appl
17	116.5	5.4	490	US-08-201-118-9	Sequence 9, Appl
18	116.5	5.4	490	US-08-238-821B-3	Sequence 3, Appl
19	116.5	5.4	490	US-08-238-821B-9	Sequence 9, Appl
20	116.5	5.4	490	PCT-US95-05744-3	Sequence 3, Appl
21	116.5	5.4	490	PCT-US95-05744-9	Sequence 9, Appl
22	116.5	5.4	501	US-08-906-791-2	Sequence 2, Appl
23	113	5.2	490	US-08-201-118-7	Sequence 7, Appl
24	112.5	5.2	490	US-08-238-821B-7	Sequence 7, Appl
25	112.5	5.2	490	PCT-US95-05744-7	Sequence 7, Appl
26	111.5	5.1	504	US-08-457-274A-25	Sequence 25, Appl
27	111.5	5.1	504	PCT-US95-05758-25	Sequence 25, Appl
28	111.5	5.1	504	PCT-US95-05758-25	Sequence 25, Appl

29	108	5.0	492	3	US-08-724-466B-2	Sequence 2, Appl
30	107	4.9	513	3	US-08-948-56A-6	Sequence 6, Appl
31	104.5	4.8	500	2	US-08-314-601-2	Sequence 2, Appl
32	104.5	4.8	500	4	PCT-US95-13051-2	Sequence 2, Appl
33	103.5	4.8	472	2	US-08-622-166A-2	Sequence 2, Appl
34	103.5	4.8	472	2	US-08-522-166A-4	Sequence 4, Appl
35	103	4.7	497	3	US-08-724-466B-4	Sequence 4, Appl
36	100.5	4.6	498	1	US-08-457-274A-24	Sequence 24, Appl
37	100.5	4.6	498	4	PCT-US95-05758-24	Sequence 24, Appl
38	99.5	4.6	490	1	US-08-201-118-1	Sequence 1, Appl
39	99.5	4.6	490	2	US-08-238-821B-1	Sequence 1, Appl
40	99.5	4.6	490	4	PCT-US95-05744-1	Sequence 1, Appl
41	98	4.5	490	1	US-08-201-118-5	Sequence 1, Appl
42	98	4.5	490	2	US-08-238-821B-5	Sequence 5, Appl
43	98	4.5	490	4	PCT-US95-05744-5	Sequence 5, Appl
44	96	4.4	490	1	US-08-201-118-11	Sequence 11, Appl
45	96	4.4	490	2	US-08-238-821B-11	Sequence 11, Appl

ALIGNMENTS

```
RESULT 1
US-09-320-878-18
; Sequence 18, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,558
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-18
```

Query Match 16.7% Score 363.5; DB 3; Length 416;  
Best Local Similarity 28.3%; Pred. No. 2,1e-29;  
Matches 97; Conservative 62; Mismatches 163; Indels 21; Gaps 7;

QY	76	EDYHFSSECPFIREGEAYDFIPTSMDDPEQOFALANOVYGMVYDKLEKRIQELA	135
DB	72	KDWR--NSTPPLTEAEALNHNMLE--DPRHRLRLKLVAREFTMRVVELLRPOVEIV	127
QY	136	CSLSESL--RPOGOCNTEDEYAEPPFIRIFMLAGLREEDIPHLKYLTDQWTRDGSMTF	193
DB	128	DGLVDALAPBDGADLMESLAWPLPTIVSELGVEBPDAARFVTDFAVFPDDPAQA	187
QY	194	AEAKAALYDVLPIPLEOROKPGTDAIS--IVANGQVNGRPTSDAKRMGLLAVGLDT	252
DB	188	QTMAKESGYLSRLIDSKRGDDGLLSALVRYTSDSDGSRITSELLGMHILLVACHET	247

```

Oy      253  VNFESEFEMEFLASPEIRCOLIERPERIPACCELLRFSIVADGILHSIDYEF-----
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      248  TVNLIANQMYALLSHPPDIALLRDMTLLDGLAVEMILR-----YEGPVEASATYFPVEPV
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      308  --HGVOLKKDQOILLPOMLSGLDERKNACPMHVDPSFOKVSHTTTFGHSILCIGOHILARR
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      303  DLDGVIPAGTGTVLVLVDHARTPEREPDFRPFIRDTGHLAFHGIGHFCIGAPLARRL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      366  EIVTLKEMLTRIPDES--IAPGAOIOHKSIGVGOALPLVW 406
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      363  EARIAVRALLERCPDLALDVSFGELWYTPMIMGALALPIRW 405
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT      2
5212296-9
PATENT NO. 5212296
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LENO, KENNETH
J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.
TEPPERMAN, JAMES M.
TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
CYTOCHROMES
NUMBER OF SEQUENCES: 19
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/569,781
FILING DATE: 23-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 464,499
FILING DATE: 12-JAN-1990
APPLICATION NUMBER: 405,605
FILING DATE: 11-SEP-1989
SEQ ID NO: 9
LENGTH: 403
5212296-9

```

[illegible]

RESULT 3  
5212296-6  
Patent No. 5212296  
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH  
J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.

```

: TEPPERMAN, JAMES M.
: TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
: CYTOCHROMES
:
: NUMBER OF SEQUENCES: 19
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/569,781
: FILING DATE: 23-AUG-1990
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: 464,499
: FILING DATE: 12-JAN-1990
: APPLICATION NUMBER: 405,605
: FILING DATE: 11-SEP-1989
:
: SEQ ID NO.:6:
: LENGTH: 406
: 5212296-6

```

[illegible]

US-08-102-863-11  
 RESULT 4  
 : Sequence 11, Application US/08102863  
 : Patent No. 5466590  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: SARLASLANT, SIMA  
 : TITLE OF INVENTION: CONSTITUTIVE  
 : TITLE OF INVENTION: EXPRESSION OF  
 : TITLE OF INVENTION: AND FERRIDOXY  
 : TITLE OF INVENTION: STREPTOMYCES  
 : NUMBER OF SEQUENCES: 12  
 :  
 : CORRESPONDENCE ADDRESS:  
 :  
 : ADDRESSEE: E. I. DU PONT DE NEMUR  
 :  
 : ADDRESSEE: AND COMPANY  
 : STREET: 1007 MARKET STREET  
 : CITY: WILMINGTON  
 : STATE: DELAWARE  
 : COUNTRY: USA  
 : ZIP: 19898  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC Compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0,
; SOFTWARE: Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/102,863
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/807,001
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: GALLEGOS, R. THOMAS
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: CR-9000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-7342
; TELEFAX: 302-892-7949
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-102-863-11
```

Query Match 14.4% Score 314; DB 1; Length 412;

Best Local Similarity 27.2% Pred. No. 2.8e-24; Matches 84; Conservative 56; Mismatches 163; Indels 6; Gaps 5;

```

OY 103 MDPEORQFRALANOVVMPVVDKLENIQELACSLIESLRPG-OCNFTDYAEPPIR 161
    :| | | | | : : : | | | | | : : : : | | |
DB 105 VDDEHNTQRMRLPTTSVKRIGALRPRIQETVRLDAMERGPRALVSAFALPVS 164
    :| | | | | : : : | | | | | : : : : | | |
OY 162 IFMLAGLPREDIDHLKYLDQMTRPDGSMTFAAKKALYDYLPIIEOROKPGTDAIS 221
    :| | | | | : : : | | | | | : : : : | | |
DB 165 VICALLGPVADHAFEEERSQRLRGPGADVNRARDELEYGALIDRKRAEPGDGLD 224
    :| | | | | : : : | | | | | : : : : | | |
OY 222 IVANGOVNGRPTSDCAKRMGILLVGLDVTVPNLFSSMEFLAKSPHRELIERPERI 281
    :| | | | | : : : | | | | | : : : : | | |
DB 225 ELIHRDHPDGVDRQOLVAFVAILLIAGHETTANMISLGFTLLSHPEOLALRAGTST 284
    :| | | | | : : : | | | | | : : : : | | |
OY 282 PAACEELLRRFSIADG--RLTSDYEFHGYOLKKGDQILLPQMSGLDEKKNACPMHYD 339
    :| | | | | : : : | | | | | : : : : | | |
DB 285 AVVVEELL-RLSLTAEGQRLATEDMEVDGATIRKGEVVESTSLINDADVPRAEITLD 343
    :| | | | | : : : | | | | | : : : : | | |
OY 340 FSRQKVSHTTFGHSHLCLGQHLARREIIVTLKEMLTRIPDFSIA-PCAQIQHKS-IVS 397
    :| | | | | : : : | | | | | : : : : | | |
DB 344 WDRPARHHLARFGVHQCIGNLARAEIDIAMRTLFEHLRGLRLAVRANHLRHKPGDTIQ 403
    :| | | | | : : : | | | | | : : : : | | |
OY 398 GVQALPLVW 406
    :| | | | | : : : | | | | | : : : : | | |
DB 404 GLDLPLVAM 412
```

RESULT 5  
PCT-US92-10885-11

```

; Sequence 11, Application PC/TUS9210885
; GENERAL INFORMATION:
; APPLICANT: SARIASLANI, SIMA
; TITLE OF INVENTION: CONSTITUTIVE
; TITLE OF INVENTION: EXPRESSION OF P450SOY
; TITLE OF INVENTION: AND PEREDOXIN-SOY IN
; TITLE OF INVENTION: STREPTOMYCES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; ADDRESSEE: AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
; ZIP: 19898
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch,
; MEDIUM TYPE: 1.0 MB
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10885
; FILING DATE: 19921216
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: GALLEGOS, R. THOMAS
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: CR-9000-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-7342
; TELEFAX: 302-892-7949
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; PCT-US92-10885-11
```

Query Match 14.4% Score 314; DB 4; Length 412;

Best Local Similarity 27.2% Pred. No. 2.8e-24; Matches 84; Conservative 56; Mismatches 163; Indels 6; Gaps 5;

```

OY 103 MDPEORQFRALANOVVMPVVDKLENIQELACSLIESLRPG-OCNFTDYAEPPIR 161
    :| | | | | : : : | | | | | : : : : | | |
DB 105 VDDEHNTQRMRLPTTSVKRIGALRPRIQETVRLDAMERGPRALVSAFALPVS 164
    :| | | | | : : : | | | | | : : : : | | |
OY 162 IFMLAGLPREDIDHLKYLDQMTRPDGSMTFAAKKALYDYLPIIEOROKPGTDAIS 221
    :| | | | | : : : | | | | | : : : : | | |
DB 165 VICALLGPVADHAFEEERSQRLRGPGADVNRARDELEYGALIDRKRAEPGDGLD 224
    :| | | | | : : : | | | | | : : : : | | |
OY 222 IVANGOVNGRPTSDCAKRMGILLVGLDVTVPNLFSSMEFLAKSPHRELIERPERI 281
    :| | | | | : : : | | | | | : : : : | | |
DB 225 ELIHRDHPDGVDRQOLVAFVAILLIAGHETTANMISLGFTLLSHPEOLALRAGTST 284
    :| | | | | : : : | | | | | : : : : | | |
OY 282 PAACEELLRRFSIADG--RLTSDYEFHGYOLKKGDQILLPQMSGLDEKKNACPMHYD 339
    :| | | | | : : : | | | | | : : : : | | |
DB 285 AVVVEELL-RLSLTAEGQRLATEDMEVDGATIRKGEVVESTSLINDADVPRAEITLD 343
    :| | | | | : : : | | | | | : : : : | | |
OY 340 FSRQKVSHTTFGHSHLCLGQHLARREIIVTLKEMLTRIPDFSIA-PCAQIQHKS-IVS 397
    :| | | | | : : : | | | | | : : : : | | |
DB 344 WDRPARHHLARFGVHQCIGNLARAEIDIAMRTLFEHLRGLRLAVRANHLRHKPGDTIQ 403
    :| | | | | : : : | | | | | : : : : | | |
OY 398 GVQALPLVW 406
    :| | | | | : : : | | | | | : : : : | | |
DB 404 GLDLPLVAM 412
```

RESULT 6  
US-09-335-409-8

```

; Sequence 8, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
```

SEQ ID NO 8  
LENGTH: 419  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-335-409-8

Query Match 12.5%; Score 271.5; DB 3; Length 419;  
Best Local Similarity 23.8%; Pred. No. 7.4e-20;  
Matches 100; Conservative 67; Mismatches 179; Indels 75; Gaps 14;

QY 6 OSNANLAPLPHVPEHVLVPEFDMNPNSLSAGVOE-----AMAVLQESNVDLVYTRCNG 61  
DB 3 QEOANQSETPK-----AFDKPPAP-----GIAEDPPALIERLEA-TPIFYWD--EGR 48  
QY 62 HWIATRGQ-----LIREAYEDYRHFSECEPIPREAGEAVDTIPTSMDPPEQR 109  
DB 49 SWVLTRHDVSAVFRDRFRAVSREMESSAYSALP-----ELSDMKKYGIFGLRPEDHA 104  
QY 110 QFRALANOVGMVYDKLENRIQLACSLIESLRPGOCNFTEDYAPFPRIIFMLIAGL 169  
DB 105 RVKRLVPSFTSRALIDLRRAEIORTVQDLIDARSGOEEFVVRDYAGIIPMAISALLKV 164  
QY 170 PEEDIRPLKTLTQMTPRDSMTFAEKEALYDILP-----I 207  
DB 165 PAE-----CDEKFRFESSAT-----ARALGVGLVPOVDEETKTVASVTEGALLHDV 212  
QY 208 IEOROKRP-GTDAISIVANGVGRPTTSDCAKRMFGLLVGIDVFNLFSEFMEFLAK 266  
DB 213 LDERRRRLPNDVITMLQAEADSGRLSTKEVALVGAITTAAGDTITTYIARVNLRLR 272  
QY 267 SPEHQELIERPERIPACELLRRFSLVADG--RIITSDYEFHGVLKKGDOI--LLPQ 322  
DB 273 SPEALELVKAEPLMRNALDEVLRFDNILRIQVRFARQDLEVCASIKKGEWVFLIPS 332  
QY 323 MLSGLDERKNACPMHVDFSQKVSHTFGSHLCLGQHARRIITLKEMLTRIDFS 382  
DB 333 ALR--DQTVSRPDVFDVRBDTGSALVGRGPHVCPSGLARLEAEIAGVITFRFPEMK 390  
QY 383 I 383  
DB 391 L 391

RESULT 7  
US-09-982-5  
Sequence 5, Application US/09096982  
Patent No. 5962293  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEE, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/096,982  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Goltick, Mary E.  
REGISTRATION NUMBER: 34829

REFERENCE/DOCKET NUMBER: 22727/00131  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 216-622-8458  
TELEFAX: 216-241-0816  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-982-5

Query Match 7.4%; Score 161; DB 2; Length 422;  
Best Local Similarity 23.6%; Pred. No. 2.2e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 QEAMVQLQESNVDLVYTRCNGHWITRQQLREAYEDYRHFSSP--CFPIPREAGEAV 96  
DB 31 REAGPVV-EVNAP-----AGRAWVITDDALAREVLADRFVKGPDLPATWAGVDDGL 83  
QY 97 D-----PIPTSMDPPEQORFALANOVGMVYDKLENRIQLACSLIESL-----R 143  
DB 84 DIVPELRFTLLAVDGEDHRLRLRIHARFNPRLAERTDRLAALADRLTLELADSSDR 143  
QY 144 POGOCNFTEDYAPFPRIIFMLIAGL-----EEDIRPLKYL-----TDQMTRPD 188  
DB 144 SGEPAELIGFAFANHPFLVYCELLGVPTPRAMAREAVGLKALGLGPGSAGDGDTPA 203  
QY 189 GSMTPFAKALYDILPIETDOROKRPGTDALISIVANGVGRPTTSDCAKRMFGLLVG 248  
DB 204 GDVPDTSALSL--LLEAVHARRKDTRTMTRVLYERAQAEFSSVSDQDLVYMITGLIFA 261  
QY 249 GLDTVFNLFSEFMEFLAKSEHQELIERPERIPACELLRR-----PSLVADGRITFS 303  
DB 262 GHDTGSFGLF-----LAEVLAGRADADGDAISREVEEALRHNPVPVSL--WRAAF 315  
QY 304 DYFHHGVLKKGQDILPQMLSGLDERKNACPMHVDFSRQKVSHTFGSHLCLGQHILA 363  
DB 316 EVVIRGVRLPRGAPVLVDIEGTMTDGRHNDARPHAFHNDRRSRRLRTFGDGRPHCIGBOILA 375  
QY 364 RREIYTLKEMLTRIDFESTA-PSAOTO--HKSGIVSGVALPLVW 406  
DB 376 QLESRTWIGVLRSPQARLAVPEELRMCCKGAGTARLTDLF-VW 420

RESULT 8  
US-08-653-650A-5  
Sequence 5, Application US/08653650A  
Patent No. 5976830  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEE, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/653,650A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:



```

: APPLICATION NUMBER: US/08/653,650A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Goltick, Mary E.
: REGISTRATION NUMBER: 34829
: REFERENCE/DOCKET NUMBER: 22727/00131
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 216-622-8458
: TELEFAX: 216-241-0816
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 474 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-653-650A-8

Query Match 7.4%; Score 161; DB 2; Length 474;
Best Local Similarity 23.6%; Pred. No. 2.6e-08;
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14:

QY 39 QEAMAVLQESNVDPDLVWTRNGGHWTATRGQLIREAEDYRHHSS--CPPIPREAGEAY 96
DB 83 REAGPVV-EVNAP-----AGGPAMVITDDALAREVLADRFYKGPDIAPTMKGVDDGL 135
QY 97 D-----FPTSMDEPREQRFALANQVGMVVDKLENRIQELACSLIESL-----R 143
DB 136 DIVPELRPTLLAVDGEDHRLRIHAPAFNPRRLAERTDRIATAIDRLTELADSSDR 135
QY 144 PQGCNFTEDYAEPPPIRIFMLAGLP-----EEDIPHLKYL-----TDQMTRPD 188
DB 196 SGPEAELIGFAHFPRLVLYCELLGVPTDPMARAVGVKALGLGPGSAGDGTDP 255
QY 189 GSMTEFAEKALDYLIPIEORRQKGTDAISIVANGVNGRPITSDKRRFGILLVG 248
DB 256 GDVPDTSALSL--LLEAVIAARRKDTRTMTRVLYERAQAEFGSVSDDLVYMITGLIFA 313
QY 249 GLDTVNVFLFSMEFLAKSPENROELIERPERIPAAACEELLRR-----FSLVADGRILTS 303
DB 314 GHDTGSFLGF--LLAEVLAGRLADADODALSRFVEELRNHPVRYSL--WKRAAT 367
QY 304 DYEPHGVLAKKGOQLLPQMLSGLDERKNACPMHVDPSRQKVSHTTFGSHLCLGQHILA 363
DB 368 EYVIRGVRLPRGAPVLDIEGTITDGRHNDAPAFHPRDRSRRLTFEGDGPYICIGOLA 427
QY 364 RREIYTLKEMLRIPDPSIA--PGAQIQ--HKSGLVSGVALPLVW 406
DB 428 QLESRTMIGVLRSPQARLAVPYEELRWCKGAGTARLTDLR-VW 472

RESULT 11
US-09-096-982-9
: Sequence 9, Application US/0906982
: Patent No. 5962293
: GENERAL INFORMATION:
: APPLICANT: Strohl, William R.
: APPLICANT: Dickens, Michael L.
: APPLICANT: Desanti, Charles L.
: TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CALFEY, HALTER & GRISWOLD
: STREET: 800 Superior Avenue, Suite 1400
: CITY: Cleveland
: STATE: Ohio
: COUNTRY: USA
: ZIP: 44114-2688
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/096,982
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Goltick, Mary E.
: REGISTRATION NUMBER: 34829
: REFERENCE/DOCKET NUMBER: 22727/00131
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 216-622-8458
: TELEFAX: 216-241-0816
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 443 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-09-096-982-9

Query Match 7.3%; Score 158; DB 2; Length 443;
Best Local Similarity 23.6%; Pred. No. 4.8e-08;
Matches 95; Conservative 50; Mismatches 210; Indels 48; Gaps 13;

QY 39 QEAMAVLQESNVDPDLVWTRNGGHWTATRGQLIREAEDYRHHSS--CPPIPREAGEAY 96
DB 52 REAGPVV-EVNAP-----AGGPAMVITDDALAREVLADRFYKGPDIAPTMKGVDDGL 104
QY 97 D-----FPTSMDEPREQRFALANQVGMVVDKLENRIQELACSLIESL-----R 143
DB 105 DIVPELRPTLLAVDGEDHRLRIHAPAFNPRRLAERTDRIATAIDRLTELADSSDR 164
QY 144 PQGCNFTEDYAEPPPIRIFMLAGLP-----EEDIPHLKYL-----TDQMTRPD 188
DB 165 SGPEAELIGFAHFPRLVLYCELLGVPTDPMARAVGVKALGLGPGSAGDGTDP 224
QY 189 GSMTEFAEKALDYLIPIEORRQKGTDAISIVANGVNGRPITSDKRRFGILLVG 248
DB 225 GDVPDTSALSL--LLEAVIAARRKDTRTMTRVLYERAQAEFGSVSDDLVYMITGLIFA 282
QY 249 GLDTVNVFLFSMEFLAKSPENROELIERPERIPAAACEELLRRFSLV--ADGRILTS DYE 306
DB 283 GHDTGSFLGF--LLAEVLAGRLADADODALSRFVEELRNHPVRYSL--WKRAAT 339
QY 307 FHGVOLKKGQQLLPQMLSGLDERKNACPMHVDPSRQKVSHTTFGSHLCLGQHILARE 366
DB 340 IRGVRLPRGAPVLDIEGTITDGRHNDAPAFHPRDRSRRLTFEGDGPYICIGOLA 399
QY 367 IIVTLKEMLRIPDPSIA--PGAQIQ--HKSGLVSGVALPLVW 406
DB 400 SRIMIGVLRSPQARLAVPYEELRWCKGAGTARLTDLR-VW 441

RESULT 12
US-08-653-650A-9
: Sequence 9, Application US/08653650A
: Patent No. 5976830
: GENERAL INFORMATION:
: APPLICANT: Strohl, William R.
: APPLICANT: Dickens, Michael L.
: APPLICANT: Desanti, Charles L.
: TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CALFEY, HALTER & GRISWOLD
: STREET: 800 Superior Avenue, Suite 1400
: CITY: Cleveland
: STATE: Ohio
: COUNTRY: USA
: ZIP: 44114-2688
: COMPUTER READABLE FORM:
```



```

? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent'n Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/653,650A
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Golfick, Mary E.
? REGISTRATION NUMBER: 34829
? REFERENCE/DOCKET NUMBER: 22727/00131
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 216-622-8458
? TELEFAX: 216-241-0816
? INFORMATION FOR SEQ ID NO: 9:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 443 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? JS-08-653-650A-9

```

Query Match	7.38;	Score 158;	DB 2;	Length 443;
Best Local Similarity	23.68;	Pred. No. 4.8e-08;		
Matches	95;	Conservative 50;	Mismatches 210;	Indels 48;
			Gaps	13

QY	39	QEANAIVLOESVWPVLVWTRCNGCMHITRCOLLREAVEDYRHFSSE--CFIIPREAGEAV	96
Db	52	REACRPV-EVNAF-----AGPAMVITTDALAREVLADRPVKDPRDLATAMRGVNDGL	104
QY	97	D-----FIPSTMDEPQOFQFALANQYVGMVYDKLENRIQELACSLIEST-----R	143
Db	105	DIPVELKRPFLVLAVDGDGHDHRRILRIHAPAFNPBRALERTPDRIALADRLTELAOSSDR	164
QY	144	POGOCNPFEDYAEFFPPIRIEKLGLP-----EEDIPHLKYL-----TDQMRPD	188
Db	165	SGEPALIELGGFAVYHFPLVLITCELLGVVPTDPMAREAVGVKLKGLGSPAGSGDGTDP	224
QY	189	GSMFPAEAKKALUYLLPIIQOROKQCTDIAISVANGQVNGRIYDEDKAKKRGILLVG	248
Db	225	GDVDPDTSLIEL--LLEAVNAARKKDRITRTMYLVEBAQAQFSGVSDDOLVUMITGLIFA	282
QY	249	GLDTPVNFNLSEMEFLAKSPREHOLEIERPERIPACEELLIRSELY--ADGRILTSDEY	306
Db	283	GHDPTTGSFLG---LLEAVLGRILAADADGDALISRFVEEALRNHPRVRYLLMPREATEV	339
QY	307	PHGVOLKKGDDILLPOMLSGLDERKNACPMHNVDSROKASHTTTFGSHGLCLOHILARE	366
Db	340	IRGVRLPRGAVLVDIEGTINTDGNHNDARNAHFNRDRPSRRILTFGGDSRPHYCIGQALQLE	399
QY	367	IIVVLKEMVILTRPDSIA-PCQAQIQ--HKSGTSGVADLPRLVW	406
Db	400	SRTMIGVLRSSRPQARLAVPVELLWRCRKGAQTARLLDLP--VW	441

RESULT 13  
US-08-396-218-2  
: Sequence 2, Application US/08396218  
: Patent No. 5695966  
: GENERAL INFORMATION:  
: APPLICANT: INVENTI, Augusto  
: APPLICANT: BREME, Umberto  
: APPLICANT: COLOMBO, Anna L  
: APPLICANT: HUTCHINSON, Charles R  
: APPLICANT: OTTEN, Sharee  
: APPLICANT: SCOTTI, Claudio  
: TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN  
: NUMBER OF SEQUENCES: 4  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: NIKAIIDO, MARNELESTEIN, MURRAY & GRAM

```

1      STREET:  655 Fifteenth Street, N. W., Suite 330 - G
2      STREET:  Street Lobby
3      CITY:    Washington
4      STATE:   DC
5      COUNTRY: USA
6      ZIP:     20005-5701
7
8      COMPUTER READABLE FORM:
9      MEDIUM TYPE: Floppy disk
10     COMPUTER:  IBM PC compatible
11     OPERATING SYSTEM: PC-DOS/MS-DOS
12     SOFTWARE:  PatentIn Release #1.0, Version #1.30
13
14     CURRENT APPLICATION DATA:
15     APPLICATION NUMBER:  US/08/396,218
16     FILING DATE:
17
18     CLASSIFICATION:  536
19
20     ATTORNEY/AGENT INFORMATION:
21     NAME:  KITTS, Monica C
22     REGISTRATION NUMBER:  36,105
23     REFERENCE/DOCKET NUMBER:  P1615-5002
24
25     TELECOMMUNICATION INFORMATION:
26     TELEPHONE:  202/638-5000
27     TELEFAX:    202/638-4810
28
29     INFORMATION FOR SEQ ID NO:  2:
30
31     SEQUENCE CHARACTERISTICS:
32     LENGTH:  422 amino acids
33     TYPE:    amino acid
34     TOPOLOGY:  linear
35
36     MOLECULE TYPE:  protein
37
38     US-08-396-218-2

```

IUPOLUGI: linear  
 MOLECULE TYPE: protein  
 S-08-396-218-2

Query Match	7.28: Score 156; DB 1; Length 422:
Best Local Similarity 23.68; Pred. No. 7.2e-08;	
Matches 97; Conservative 51; Mismatches 202; Indels 58; Gaps	
QY	QENAAVLIQESVNPDLVWTRCNGGMIATRGOLIAREAYEDYRHNSESSECPITPREAGEAYDF 98
Db	31 REAGVIV- EVNAP-----ACGPAMVITLDALAREVLADPRFVVD-----PDLAPAMRG 78
QY	99 IPTSMD--PEPQKQFRLA-----NQYVGMVVD-----KIENIQLQSLIESL- 142
Db	79 VDDGDIPLVPELRPEFTLIAVDGAAHRRILRIHAFAPNPRRIAEETDRIAAIAGRLITELA 138
QY	143 -----RPOQGCNFTEDYAEPPRIPIRFLMLAGLP-----BEDIPHLKYL-----TDQ 183
Db	139 DASGRSGKRAELIGGFAYHNPFLVITCYELLGVPYTDPPMAKEAVSVLALGLGCPQSGGD 188
QY	184 MTRPDGSMTEFAEAKALYDVLPIPIEOROKPGTDAISIVANGOVNRPITTSDEAKRMFC 243
Db	199 GTDPAGGAGPDTSALESIL--LLEAVHSAIRNDTPMTVFLYERAAQAEFGSVDDQLVYMIT 256
QY	244 LLIVGGIDTVVNFISFMSHEFLAKSPREHQLDIERRERIPAAACEFLIRFSLV--ADSRIL 300
Db	257 GLIFRAGHDTTGSFICG---LTAELVLAGLADADADADADVSRFVEALYINRPVPYTLMRFA 313
QY	302 TSDYEFHGVQLKKGDQILLPQMLSGIDERKNACSPYHNVDFSQOKVSHTFGHSNLCIGH 361
Db	314 ATEVATIGVRLPRGARVYLVDIEGTNTDGRHNHDARHAFRDRPWSMRRLTFGDPHYICGEQ 373
QY	362 LAREEIVTLKEMUIRIPDPSIA-PMQAOIQ--HKSGLVSGVQALRPLW 406
Db	374 LAQLESRTMIGVLSRPEARALAYUDELKRCRGAQOTARLTIRL- VW 420

RESULT 14  
US-08-760-116-2  
: Sequence 2, Application US/0876016  
: Patent No. 5766190  
: GENERAL INFORMATION:  
: APPLICANT: INVENTI, Augusto  
: APPLICANT: BREME, Umberto  
: APPLICANT: COLOMBO, Anna L  
: APPLICANT: HUTCHINSON, Charles R



Db 365 PYMEAFILLETFRRHSSFVPTIPNSTRDTSLKGFYIPKGRGVFNQMOINHDOKLWVNP 424  
QY 336 -----MHVDFSROKY--SHTTEGHSHLCLGOHLARREIIVTLKENLTRLRIDPFSIAP 385  
Db 425 EFLPERFLTPDGAIDKVLSEKVIIFGMRKRCIGETVARWEVFLFLAILLQRY-EFSVPL 483  
QY 386 GAQI 389  
Db 484 GVKV 487

Search completed: October 4, 2000, 12:38:09  
Job time: 21426 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:36:43 ; Search time 73.09 Seconds  
(without alignments)  
350.531 Million cell updates/sec

Title: US-09-246-451-13  
Perfect score: 2177  
Sequence: 1 TTTETIQSNANLAPLPHPVPE.....IVSGVQALPLWDPATTKAV 414

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR\_64:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2165	99.4	415	1	O4PSCP
2	376	17.3	393	1	C70829
3	371	17.0	410	2	S39924
4	362	16.6	397	1	S51594
5	359	16.5	398	1	H70807
6	355	16.3	396	1	B69851
7	346.5	15.9	428	1	E70729
8	344.5	15.8	404	2	T30231
9	340	15.6	403	2	B35401
10	337	15.5	411	1	JC5859
11	334.5	15.4	410	1	O4B86M
12	334	15.3	410	1	E69611
13	333.5	15.3	405	1	B42606
14	328	15.1	406	2	A35401
15	327	15.0	406	1	A48495
16	325.5	15.0	376	1	G69679
17	323	14.8	428	1	A42971
18	318.5	14.6	405	1	H70752
19	317.5	14.6	399	1	A55578
20	314	14.4	412	2	S24750
21	311.5	14.3	410	2	JC4287
22	307	14.1	310	2	T44767
23	305.5	14.0	395	1	G69594
24	303	13.9	417	1	S49051
25	302	13.9	438	1	E70515
26	301	13.8	402	2	JC5151
27	301	13.8	404	1	JC5150
28	294.5	13.5	408	2	PD0007
29	294.5	13.5	408	2	JC5674

30	291.5	13.4	406	1	S18531	cytochrome P450 er
31	291.5	13.4	433	1	B70677	cytochrome P450 rv
32	285	13.1	381	1	S15809	cytochrome P450 cy
33	283.5	13.0	337	2	T30601	cytochrome P450 hy
34	273	12.5	414	1	E70708	cytochrome P450 -
35	269.5	12.4	386	2	T30235	cytochrome P450 -
36	266.5	12.2	310	2	T44857	probable hydroxyla
37	265	12.2	412	1	B40634	erythromycin mono
38	251.5	11.6	401	2	T36526	probable cytochrom
39	251	11.5	401	1	I40208	cytochrome P450 bt
40	250	11.5	411	1	JC4003	cytochrome P450 -
41	244	11.2	406	3	T17487	cytochrome P450 hy
42	240	11.0	433	2	T44587	cytochrome P450 ho
43	237	10.9	402	1	A70707	cytochrome P450 rv
44	235	10.8	396	1	H70730	cytochrome P450 rv
45	225	10.3	398	1	S47520	vitamin D-3 25-hyd

## ALIGNMENTS

RESULT 1  
O4PSCP  
camphor 5-monooxygenase (EC 1.14.15.1) cytochrome P450 101 - Pseudomonas putida plasm  
N:Alternate names: cytochrome P450-CAM  
C:Species: Pseudomonas putida  
C:Date: 30-Apr-1982 #sequence.revision 31-Dec-1993 #text\_change 03-Mar-2000  
C:Accession: A25660; S34614; C60886; A00194  
R:Unger, B.P.; Gunsalus, I.C.; Sligar, S.G.  
J. Biol. Chem. 261, 1158-1163, 1986  
A:Title: Nucleotide sequence of the Pseudomonas putida cytochrome P-450-cam gene and  
A:Reference number: A94678; MUID:86111751  
A:Accession: A25660  
A:Molecule type: DNA  
A:Residues: 1-415 <DNC>  
A:Cross-references: GB:M12546; NID:g151114; PIDN:AAA25760.1; PID:g151115  
R:Atamaki, H.; Koga, H.; Sagara, Y.; Hosoi, M.; Horuchi, T.  
Biochim. Biophys. Acta 1174, 91-94, 1993  
A:Title: Complete nucleotide sequence of the 5'-exo-hydroxycamphor dehydrogenase gene  
A:Reference number: S34613; MUID:93326643  
A:Accession: S34614  
A>Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-42 <ARA>  
A:Experimental source: PpG1; ATCC 17453; CAM plasmid  
R:Romeo, C.; Moriwaki, N.; Yasunobu, K.T.; Gunsalus, I.C.; Koga, H.  
J. Protein Chem. 6, 253-261, 1987  
A:Title: Identification of the coding region for the putidaredoxin reductase gene fro  
A:Reference number: A60886  
A:Accession: C60886  
A:Molecule type: DNA  
A:Residues: 408-415 <ROM>  
R:Hanlu, M.; Armes, L.G.; Yasunobu, K.T.; Shastri, B.A.; Gunsalus, I.C.  
J. Biol. Chem. 257, 12664-12671, 1982  
A:Title: Amino acid sequence of the Pseudomonas putida cytochrome P-450. II. Cyanogen  
A:Reference number: A00194; MUID:83030788  
A:Accession: A00194  
A:Molecule type: protein  
A:Residues: 2-55, 58-276, 'Q', 278-361, 'S', 363-407, 'N', 409-415 <HAN>  
C:Genetics:  
A:Gene: camC; CYP101  
A:Genome: plasmid  
C:Function:  
A:Description: catalyzes hydroxylation of camphor to yield 5-exo-hydroxycamphor; elec  
C:Superfamily: Pseudomonas plasmid camphor 5-monooxygenase; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygena  
F:246-380/Domain: cytochrome P450 homology <CYP>  
F:358/Binding site: heme iron (Cys) (axial ligand) #status experimental

Query Match 99.4%; Score 2165; DB 1; Length 415;  
Best Local Similarity 99.5%; Pred. No. 3.6e-157;  
Matches 412; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY	1	TTTTTQSANLAPLPHPHPEHLVEFDMYNPSNLISAGVOEAMAYLOESNVDDLWTWTRNG	60
Db	2	TTTTTQSANLAPLPHPHPEHLVEFDMYNPSNLISAGVOEAMAYLOESNVDDLWTWTRNG	61
QY	61	GHMIAIATROQLIREAEEDYRHFSSSECPFIPIREAGEAYDIPISMPDPEORQIPALANOVG	120
Db	62	GHMIAIATROQLIREAEEDYRHFSSSECPFIPIREAGEAYDIPISMPDPEORQIPALANOVG	121
QY	121	MPVVDKLENRIQOELACSLIESIRPOGOCNFTEDYAEPPRIIFMLLAGLPEDIDPHLYL	180
Db	122	MPVVDKLENRIQOELACSLIESIRPOGOCNFTEDYAEPPRIIFMLLAGLPEDIDPHLYL	181
QY	181	TDQMRPRGSMTPFAAKALYDYLPIITEORRKPQTALISIVANGOVNGBRPTISDEKKR	240
Db	182	TDQMRPRGSMTPFAAKALYDYLPIITEORRKPQTALISIVANGOVNGBRPTISDEKKR	241
QY	241	MFGLLVAGGLDITVNVFLSFSMEFLAKSPENHOELLERPERIPACEEYLIRFSTIVADGRI	300
Db	242	MCGLLVAGGLDITVNVFLSFSMEFLAKSPENHOELLERPERIPACEEYLIRFSTIVADGRI	301
QY	301	LTSDEYFHGVOLKKSDQILLPQMLSGDLDERKNACPMHYDFSROKVSHTTFGHGSHLCIG	360
Db	302	LTSDEYFHGVOLKKSDQILLPQMLSGDLDERKNACPMHYDFSROKVSHTTFGHGSHLCIG	361
QY	361	HLAREELIVTAKEMLTRIPDPSIAPGAOIOHKSIGVSGVQALPLVMPDATTKAV	414
Db	362	HLAREELIVTAKEMLTRIPDPSIAPGAOIOHKSIGVSGVQALPLVMPDATTKAV	415

RESULT 2  
C70929  
cytochrome P450 RV1785c - Mycobacterium tuberculosis (strain H37RV)  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Mycobacterium tuberculosis  
C:date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: C70929  
R:Coile, S.T.; Brosch, R.; Parthill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
A:Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: C70929  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-393 <COI>  
A:Cross-references: GB:AL022021; GB:AL123456; NID:3250699; PTDN:CAA17707.1; PID:e125459  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV1785c  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase

Query Match	17.3%	Score 376;	DB 1;	Length 393;
Best Local Similarity	28.7%	Pred. No. 3.8e-21;		
Matches 110; Conservative	70;	Mismatches 183;	Indels 20;	Gaps 10;

[illegible]

Db	194	MVAERRNPRGPDVLISQV--QIGEDPLSEIEVLGSLHLLLAGIDPVTAAVGRSLEIELAR	250
Qy	267	SPENRQELIERPEETIPAAECIELLR-RSLVADGRITLTSDEFFNGVQLKKEDQILLEFOMLS	325
Db	251	RPOLRAMLRDMPKQIRVFETIEIVLELSPARVAPRVTTPEVTVGSMTPACSPVRLCMAAV	310
Qy	326	GIDERKNACPMHVPFSROKVSHTFFGSHILCSLOHILAREEIIIVTLKEWITRTLPDESIAF	385
Db	311	NRDSDAMSTDEIVMDGKGVNHNHMGFGGGPRHRCGSHLARLELLTLVGEWLNQIPDEELAR	370
Qy	386	--GAQIQHKSGTSGVQALPLVW	406
Db	371	DYAEIIRPPSKSFA-LKNLPLRW	392

RESULT 3  
S39924  
cytochrome P450meg - *Bacillus megaterium* (ATCC 13368)  
C:Species: *Bacillus megaterium*  
A:Variety: ATCC 13368  
C:Date: 07-Oct-1994 #sequence\_revision 23-Feb-1996 #text\_change 04-Mar-2000  
C:Accession: S39924 #S32216  
R:Rauschenbach, R.; Isernhagen, M.; Noeske-Jungblut, C.; Boisdol, W.; Stewert, G.  
MOL. Gen. Genet. 241, 170-176, 1993  
A:Title: Cloning sequencing and expression of the gene for cytochrome P450meg, the str  
A:Reference number: S39924; MUID:94045677  
A:Accession: S39924  
A:Molecule type: DNA  
A:Residues: 1-410 <RAU>  
A:Cross-references: EMBL:Z21972; NID:g288298; PIDN:CA479985.1; PID:g288300  
A:Experimental source: ATCC 13368  
C:Superfamily: *Bacillus* cytochrome P450 CYP106: cytochrome P450 homolog  
C:Keywords: chromoprotein; heme; iron; metalloprotein  
C:Keywords: chromoprotein; heme; iron; metalloprotein  
C:Binding site: heme iron (Cys) (axial ligand) #status predicted  
C:Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	17.0%	Score 371,	DB 2,	Length 410,
Best Local Similarity	30.1%	Pred. No. 9.6e-21,		
Matches 109, Conservative	53,	Mismatches 180,	Indels 20,	Gaps 7

OY	63	MIITRGGLIREAYEDYHNHSESE-----CPPIREAGEAAYDTP-PSMDRRERORALAL	115
		:::	:::
Db	44	WNVKVEYKAVVLSDIKNFFSVKRTTITSVTGBSEESVEKQITPESDPRDHNKRSLL	103
		:::	:::
OY	116	NOVVMGMPVYDKLENRIQELACLSTLESRLROQSCNFTEDUAREPPIRIFMLAGYREDIP	175
		:::	:::
Db	104	AAAFPRSLQWBERIPIADELIGOMDIMGSEIDIVASLSPIRTIYAMALMSPVSDRY	163
		:::	:::
OY	176	HKLKLIDQMTGRDQSMIFAE-----AKELAYULYPIIIEOROKRQSTADISIVANGOV	228
		:::	:::
Db	164	LFKKMYDILFLPRFBREKOEVDKIKOUAAKAYUYUULPRIVQKRYNADDIISDLKSEY	223
		:::	:::
OY	229	NGRPITSDKANRMFGLLVLGSLDTVNVFLSPM-EFLAKSPREHQELLERRERIPAASEE	287
		:::	:::
Db	224	DGEMTDTDEVVTRTMLILGAGVETTSNLANSPYLLIYDQKEVUQELNEMIDLPOAVEE	283
		:::	:::
OY	288	LLR-RESLVADGRILTSDYEFNGVQOLKKGOIILRYQLSGIDREKKNASCYNVDEFSR-QKV	345
		:::	:::
Db	284	MIRREFNLIKIDRYKEDNDLGLVELKEGDSVUVMMSAAMMDVEMGERDPTLLIYNRNK	343
		:::	:::
OY	346	SHTFEGHGSNLSGLONLARREIIVTLKEWLTIRIPDFSIAFGAOLQNK--SGIVSGVAL	402
		:::	:::
Db	344	KHLTFGNGPRHPCLAGRLARLEAKTALTALFLKFKKHNIAVPSFOLEENMLDSAGOTLTSL	403
		:::	:::
OY	403	PL 404	
Db	404	PL 405	
RESULT	4		
S51594			
ytocchrome P450 mycG - Micromonospora griseorubida			

```

C:Gene: RV3518c
A:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: oxidoreductase

Query Match      16.5%; Score 359; DB 1; Length 398;
Best Local Similarity 28.2%; Pred. No. 7.6e-20;
Matches 107; Conservative 69; Mismatches 169; Indels 34; Gaps 8;

QY      55  WTRGN-----GGHWLITRGQLIRGAEVDYRHSS-----ECPTIPREAGEAVDPIPT 101
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      26  WMRANQVYFDRNRGLAASTYQAVIDAERQPELEFSNMAGIRPDPAIP-----MMI 76
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      102  SMDPEQRQFALANQVGVGVPVDKLENRIQELACSLIESLRPGQCFTEDVAEPPPIR 161
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      77  DMDPAHLIRKKILVNAAGFTRRKRVKDKKASIAALCDTLIDAVCEGSGECDFVRLAAPLMA 136
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 137 VIGDMLGVRRPQRMFLRMSDVLVTFSSHVQSQEPQITM-DATAAINDPFRATIAARRA 195

Qy 214 KPGDAISIVANGVGNRPRTISDEAKRMFGLLVGLDVTVPNLFSPFMEFLAKSPERHOE 273

Db 196 DPTDDIVSYLVSSVEDEERLSDDELVMETLLILGDETFTHRTLSGGTEQLRRRDQMDL 255

Qy 274 LIERPERIPAAACEELLRFSLVAD-GRILTSDEYEPFHVOLKGDQOILLPOMLSGLDERKN 332

Db 256 LQRPDSLLPGAIIEEMLRWATPAVKNMCRVLFTADTEFHGALCAGBKMLLFESANFEAVF 315

Qy 333 ACPHNVHPSRQKVSHTTFGHSILCLGQHLARREIIVTLKMLTRIPDFSTAPCAQI--Q 390

Db 316 CEPKPFVQVRPNPNHLLAFGFGTHFLCIGNOLARLELSLMTERRLRRLDPLRVADSVLPL 375

Qy 391 HKSGIVSGVALPLVMDPA 409

Db 376 RPAFVSGLESPVPTPTPS 394

RESULT 6

B69851

Cytochrome P450 yJb - *Bacillus subtilis*

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: *Bacillus subtilis*

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: B69851

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berni, C.; Bron, S.; Bruggliet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; C.; Enhllich, S.D.; Emerson, P.T.; Entlian, K.D.; Ertlington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galiech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Mathers, P.; Wipert, A.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Moriga, K.; Ogawa, K.; Ogilwa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetere, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scallan, A.; Authors: Schleich, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Seakeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Wilthers, Yoshikawa, H.F.; Zunststein, E.; Yoshikawa, H.; Dancho, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*

A:Reference number: A69580; MUID:98044033

A:Accession: B69851

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1396 <KUN>

A:Cross-references: GB:299110; GB:AL009126; NID:92633472; PIDN:CAB13078.1; PID:e118322

A:Experimental source: strain 168

C:Genetics:

A:Gene: yJb

C:Superfamily: *Bacillus cytochrome P450 CYP106*; cytochrome P450 homology

C:Keywords: oxidoreductase

C:236-371/Domain: cytochrome P450 homology <CYP>

```

Query March 16.3%; Score 355; DB 1; Length 396;
Best Local Similarity 30.1%; Pred. No. 1.5e-19;
Matches 99; Conservative 54; Mismatches 158; Indels 18; Gaps 7;

QY 71 IREAVEDRHFSSECPPIPREAGEAYDFIPTMDPEOROFALANQVGMVYDKLENR 130
   ::: | ||| |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 56 VKKVVGDKEFLPS-C-MPQOTSIGNSI-INMDPRKTKITSYVKKAFPTPRVMQWEDR 111
   |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
QY 131 IOELACSLIESLRPGQGNFTEDYAEPPPIRIEMLLAGPREDIPILKYLDQMT---- 185
   ||| |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 112 IOETDELIOKFOGSEFPDLYVDFSYPLVIAISELLGVPASAHMQFRAMSDLVSTPKD 171
   ||| |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
QY 186 -RPGDSMTFAE----AKKALDYLIPIIEQRQRKGTALISVANGQVNGRPITSDKKR 240
   |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 172 KSESEKAFLEERDKCEBELAAFFAGITIEEKNRKEODIISLVBABETGELSGEELIP 231
   ||| |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
QY 241 MEGLLVAGLDLVVNFILFSMEFLAKSPENHOELERPERIPAAECELLR-RESLVADGR 299
   |||| |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 232 FCTLLLVAGNETTTNMLISNAMYSLIETFGVYIELRSHPELMQAVEELRRPARPVLYLR 291
   |||| |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
QY 300 ILTSDYEFHGVOLKKGDQDILLPQMLSGIDERNKACPMHVDFSROKVSHTTGHGSHCLIG 359
   |||| |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 292 IAKRPDTEIGHILIKGCDWVLAFAVGSANRDEAKFDRPHMFDIRRHNPRIALCHGHIFCTG 351
   |||| |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
QY 360 QHLARRELIIVLKEMLTRIPDF---SLAP 385
   ||| |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 352 APLARLEANIALTSLISAFPHMECVSITP 380
   ||| |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |

```

RESULT 7  
F70729  
cytochrome P450 Rv2266 - Mycobacterium tuberculosis (strain H37Rv)  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Mycobacterium tuberculosis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: F70729  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fieldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Accession number: A70500; MUID:96295987  
A:Accession: F70729  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-428 <COL>  
A:Cross-references: GB:Z77163; GB:AL123456; NID:g3261610; PIDN:CAB00969.1; PID:e255175;  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: Rv2266  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase  
C:264-401/Domain: cytochrome P450 homology <CYP>

	Query Match	15.9%	Score 346.5;	DB 1;	Length 428;
	Best Local Similarity	26.5%;	Pred. No. 7.5e-19;		
	Matches 118; Conservative	72;	Mismatches 187;	Indels 69;	Gaps 17;
Oy	IQSNANLAPLPHPHPEHLVFPEDMYNPENLSAGVOEAMAVQOE-----SNVPDLV 54           Db	IATRNNGRP-PPEPV---IADIEL-----GSLDFPALDDVDYRDCAFLRLREAPISF 54 			
Oy	5 WTR-----CNGSHWIATRGLIREAIEDY--RH---ESS-----ECPTIPREA 92           Db	WPTTELPGFVGANGHMALTK-----YDVFVASRHPDISSPNITINDQTELAIEYF 107 			
Oy	93 GEADVDTFSMDPPEQRQFRALANQVMGPVVDKLENIQLACSLTSL---RPOGCN 149           Db	108 GSM-----IVLDPPHQHKLRISVRSAFTKPKVARLEAAVRRAHRLTVSSMANNPDRCAD 162           : : : : :   : : : : :			

QY 150 FTEDVAEEEPRIREFLLAGLPEEDIPLKYLTDQMT---RDGSGMTAFKAEKALYD---Y 203  
Db 163 LVSELAGPLPILICDDMMGIFRADRRIETHMTNVLFGFDGLLTDDEEFMVSADIGAY 222  
QY 204 LPILEORRORPGTDAISIVANGOVNGRPISDEAKRFRGGLLVGLGDLTVNLFSPSMEF 263  
Db 223 ATALAEEDRVNHHDDLTSSLVEAEVDGERSLSREIASPFILLVAGNETTBNALITHGYLA 282  
QY 264 LAKSPENHOELLERERT-PACCELLRRFS-LVADGILTSYDEPHGVQOLKKDDQLLP 321  
Db 283 LSRPEQGDWRMWSDFDGLAPFAVEEIVMVASPVYMRRTLLQDLELKGTKMAADKYSIW 342  
QY 322 QMLSGIDREKNACPMHNVFESROKVSHTTF--RHGSHLCTGOLARREIIVTLKEWLTIRPD 380  
Db 343 YCSANRDSKFAADPTPFLAARNPNHILFGGGGAHFECGLANLARREIRVAFDELRRQMPD 402  
QY 381 FSIAPGAQIOHKSGIVSGVQALPLYW 406  
Db 403 V-VATEEPARLLSQFIHGIKTLPTW 427

RESULT 8  
T30231  
cytochrome P450 - Streptomyces hygroscopicus  
C:Species: Streptomyces hygroscopicus  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: J30231  
R:Rapaporto, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; St  
Gene 169, 9-16, 1996  
A:Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces  
A:Reference number: 220782  
A:Accession: T30231  
A:Status: preliminary; translated from GB/EMBL/DBD3  
A:Molecule type: DNA  
A:Residues: 1-404 <APA>  
A:Cross-references: EMBL:X86780; NID:9987088; PID:9987105; PIDN:CAA60465.1  
C:Genetics:  
A:Gene: rapN  
A:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

Query Match	15.8%	Score 344.5	DB 2	Length 404
Best Local Similarity	28.6%	Pred. No. 9.8e-19		
Matches 102	Conservative 63	Mismatches 177	Indels 15	Gaps 9
QY	63	WATFGQLIRAYEDYRHFSECEPF--IPMEAGEAYDF-IPMSDPPEOROFALANOV	118	
Db	50	WLAASWEDVAVVFYDPR-FRSRATLGNKVPRVLAIDQPYIMDPEPHTRLRVARATKA	108	
QY	119	VGMFVVDKLEKRIQELACLLESIRPQG-OCNFEDYAEPPRIKIFMLAGLPEEDIPHL	177	
Db	109	LTSRRMEALRRTQTEVADDLIDKMLAGAPADIMEDALPLPIIMICELLGVPTEIDQTKF	168	
QY	178	KYLTPDMTRPPGSMT---FAEAKELAYDYLLPIIEOROKGTDAISIVANGVNGRPIT	234	
Db	169	RTWSQOML-SNGATSGQEVMAAGSLTYLYSELAEKRRKQDTNDLGLSVARAKKDDRLS	227	
QY	235	SDEAKRMFGLLVGLDPTVNFLEFSMEFLAKSPDHQEOELIERERIIPACCELLRRFSL	294	
Db	228	ETELVGFAYTLLINGVEETTANAGSNVYTLTLTHPEKLAELRKDLSLIPKAVDELLRIPI	267	
QY	295	VADG---RILTSDFEFHGVOLKGDQILLPOMLSGLDERKNACPMHNDSEKRVSHTFG	351	
Db	288	AKQASWVMAYEDELSEGTIVKAGEAVALIQHSANTPCKVYDHEELIDFHTSNPHSLIG	347	
QY	352	HGSHLICLOHLARREIITVTLKEWLTIRPDFSIA-PGAQIOHKSG-IYSQOALPLVW	406	
Db	348	HGAHHCMGAQVLREVMQTALGSLISRTPLRALFRANPPRRINFLKRLVPLSLEALPLTW	404	
RESULT	9			
	335401			



cytochrome P450105B1 ~ Streptomyces griseolus  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Streptomyces griseolus  
C:Date: 31-Mar-1992 #sequence,revision 31-Mar-1992 #text,change 04-Mar-2000  
C:Accession: B35401  
R:Omer, C.A.; Lestrera, R.; Little, P.J.; Dean, C.; Tepperman, J.M.; Leto, K.J.; Romesser, J.R.; Bacteriol. 172, 3335-3345, 1990  
A:Title: Genes for two herbicide-inducible cytochromes P-450 from Streptomyces griseolus  
A:Reference number: A35401; MUID:90264332  
A:Accession: B35401  
A:Molecule type: DNA  
A:Residues: 1,403 <OME>  
A:Cross-references: GB:M32239, NID:g153480, PIDN:AAA26825.1, PID:g153481, GB:M36481  
C:Genetics:  
A:Gene: CYP105B1  
A:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;  
E:28-374/domain: cytochrome P450 homology <P45>  
E:352/blinding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	15.6%	Score	340	DB 2	Length	403			
Best Local Similarity	28.9%	Pred. No.	2	1e-18					
Matches	105	Conservative	63	Mismatches	165	Indels	30	Gaps	13

QY	63	WIAIRGOLIREAYEDYRHFSECC-----PIPREAGAYDIFIPT--SMOPEEROFAPALA	115
Db	52	WLATRRHODVRAVLCDPR-FSDADHARTGFPFLTAGCGRICTGNPEFLRMDDPEHARLRRL	110
QY	116	NOVVMGPMVADKLENRIOELACSLIESLR-OGOCNFTEDYAEFPPIIFEMLAGLPEEDI	174
Db	111	TADFIIVKKEAMREYORLADDDVDRMTGTGSADLTETAFRLPLSLVLTLLGLVPEDH	170
QY	175	PHL-----KYLTDMTRPDGSMTFEAKKALYDYLPIIEBOROKPGTDAIS-IVANGVN	229
Db	171	AFEOERSVLLTTLLESTPE---EVRNADDELLEYARLARARKREPPDAIISRILVARELD	227
QY	230	GRPTISDEAKRMFELLLVGLGIDTVVNLSPSMEFKASPEHQLELPERPERIACACELL	289
Db	228	DTQIAT-----MGRLLVAGHETTANNTALSTVLVLRNPDLARLRAPALVGAEBELL	282
QY	290	RRESIVADG--RIITSDYEFHGVOLKKGDOILLPOMLSGDERKNACP--MHYDFSRQV	345
Db	283	RYLTIYHNGVPRINATEDVLLIGRTIAGEGVLC--MISSANRDAEYVPGGDDDLVDVARDAR	340
QY	346	SHTTFGHSHLCLGOHLARREIITYLKEMLTIRIPDSIA-PGAOIHKSGI-VSGVOALP	403
Db	341	RHVAFGFGVHOCLOPLARVELQIAIETLLRLRPDLRLAVPHEEIPFRGDMATIGVHSLP	400
QY	404	LVM 406	
Db	401	IAW 403	

RESULT 10  
JC5859  
polyketide synthase cytochrome P450 chain 10 - Actinomadura hibisca  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Actinomadura hibisca  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: JC5859  
R:Daici, T.; Hamano, Y.; Igarashi, Y.; Furumai, T.; Oki, T.  
Biosci. Biotechnol. Biochem. 61, 1445-1453, 1997  
A:Title: Cloning and nucleotide sequence of the putative polyketide synthase genes for F  
A:Reference number: JC5850, MUID:97480928  
A:Accession: JC5859  
A:Molecule type: DNA  
A:Residues: 1-411 <DAI>  
A:Cross-references: DDBJ:DB7924; NID:g2580441; PTDN:BAA23153.1; PID:d1024029; PID:g2580441  
C:Comment: This enzyme is involved in prodmtcin A biosynthesis.  
C:Genetics:  
A:Gene: pms10  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

**C; Keywords:** antibiotic biosynthesis; oxidoreductase F; 245-382/Domain: cytochrome P450 homology <CYP>

Query Match	15.5%	Score 337	DB 1	length 411
Best Local Similarity	27.0%	Pred. No.	3.7e-18	
Matches 117	Conservative 60	Mismatches 196	Indels 60	Gaps 17

```

OY      4  TIOSNANLAPRPPIVPEHILVDFPMYNP-----SNLSAGVOEAMVLOESNVPDLWTRCN  59
Db      9  TVDRPDPVTPAFPRRPD-----DPFQRPCEHARLKRSDPAKVLP-----T  50
OY      60  GGH-WIATRGOLIREAVEDYRHESSEC---PIIPREAGEAVDTPTS-----MDPE  107
Db      51  GDHAMVUTRYADVFVTSI-D-RFSKEAVTRGCAR-----LLPMORGSKSLVIMDPE  102
OY      108  QORFALANOVGMPPVVDKLENRLOELACSLIESLRPG-OCNTEDEYAPPEPRIFMML  166
Db      103  HTRMKATVSRAPFTARRVEGMANHRDLSGVEDMEVEHGPAPDLIAHLPLPTVTCM  162
OY      167  AGLEPEEDIPHLKYLTDQMTRPDGSMTFAEANE-----ALYDILPIIEOROKPCTDAI  220
Db      163  LGVPEDERPRRDQWTDRL-TIGAPALAOADEIKAAVGRGLYLAELIDAKTAPADDL  221
OY      221  SIVANGOVNRPITSDAEKRMFGL-LVYGLDIYVNFLLSSMEFLAKSPENROELIERPE  279
Db      222  SILSNAHADDS-LSEBELTFMGWTLAAGYHNTTAAITHSVYHLIREPSYARLRDPS  279
OY      280  RIPACCELLERFSLVAGS-----RLTSDYEFHGVOLAKGQDILLPQMLSGDERKNACP  335
Db      280  GIPAAVEELL-RYQOIGGAGACITAAVEDVEGGLTVRAGEAVIPLFNAAANRDEPVFADP  338
OY      336  MHVDFSRQVSHTTGCHGSHLICLOHARREIIVTLKEMWLRIPIDFSIA-PGAQIOHSG  394
Db      339  EELDIGRTDNPHIALHGCHYHCLAPARLEIQAUVLETVLERTPALRIDADADITWRPG  398
OY      395  IV-SGVQALPLVM  406
Db      399  LAFARPDALPTAW  411

```

RESULT 11  
O4BS6M  
cytochrome P450 106 - *Bacillus megaterium*  
N:Alternate names: cytochrome P450BM-1  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: *Bacillus megaterium*  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 03-Mar-2000  
C:Accession: S07764; S17973  
R:He, J. S.; Ruettlinger, R. T.; Liu, H. M.; Fulco, A. J.  
Biochim. Biophys. Acta 1009, 301-303, 1989  
A>Title: Molecular cloning, coding nucleotides and the deduced amino acid sequence of  
A:Reference number: S07764; MUID:90089408  
A:Accession: S07764  
A:Molecule type: DNA  
A:Residues: 1-410 <HEU1>  
A:Cross-references: EMBL:X16610; NID:g39626; PIDD:CMA4612.1; PID:g39627  
A:Accession: S17973  
A:Molecule type: protein  
A:Residues: 1-25 <HEJ2>  
C:Genetics:  
A:Gene: CYP106  
C:Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase  
F:1-110/Product: cytochrome P450 106 #status experimental <MAT>  
F:241-378/Domain: cytochrome P450 homology <CYP>  
F:356/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	15.48;	Score 334.5;	DB 1,	Length 410;
Best Local Similarity	27.78;	Pred. No. 5.7e-18;		
Matches 96;	Conservative 57;	Mismatches 170;	Indels 23;	Gaps 7.



[illegible]

RESULT 14  
A35401  
cytochrome P450 105A1 - Streptomyces griseolus  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Streptomyces griseolus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 04-Mar-2000  
C:Accession: A35401  
R:Omer, C.A.; Lestira, R.; Little, P.J.; Dean, C.; Tepperman, J.M.; Leto, K.J.; Romesser, J.  
J. Bacteriol. 172, 3335-3345, 1990  
A:Title: Genes for two herbicide-inducible cytochromes P-450 from Streptomyces griseolus  
A:Reference number: A35401, MUID:50264332  
A:Accession: A35401  
A:Molecule type: DNA  
A:Residues: 1-406 <OME>  
A:Cross-references: GB:M32238; NID:g153477; PIDN:AAA26823.1; PID:g153478; GB:M36480  
C:Genetics:  
A:Gene: CYP105A1  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;  
F:241-377/domain: cytochrome P450 homology <P45>  
F:355/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	15.1%	Score 338	DB 2	Length 406
Best Local Similarity	26.3%	Pred. No. 1.8e-17		
Matches 104	Conservative 74	Mismatches 193	Indels 24	Gaps 11

```

OY 31 PSNLSAGVO--EAAVAIOESVPLVWTRCNGH-WIATQIGOLIAREVEYU----- 79
Db 17 PSNNSCPQLPDGYAQLRDTGPRILHRTLYLXGKAMAVYIHNHARKLLGDPRLSSNRTDD 76
OY 80 HFSECEPIT--REAGEADYDFIPRSMDBPEOROFALANQVGMPPVYDKLENNIOIACS 137
Db 77 NEPATSPRFEAVRESPOAF----IGLDRPEHCTRRRTTISEFPYKRIKGRNPEVEEYVHG 132
OY 138 LIESLRPOG-QCNTEDYAEFFIRITMLLAGREEDIRPLKYLTDQMTPRDSMTPEA 196
Db 133 FLDEMLAGPRAADIVSOFALRPVSMVICRILCYADHNEFPOASKRIVOSTQASALTA 192
OY 197 KEALYDLIPIIEOROKRPGDAI-SIVANQVNGRITPTEDEAKRMGCLLVGGLDTVYN 255
Db 193 RNDLAGYLDGLTDFQYREPGALVGAALVAOQNGE-IDREELLSTAMLLIGHETPAS 251
OY 256 FLFSMEELIAPSPHEBOELIERPERIPACBELLRRESL--VADGRITLDEYHNGYOLK 313
Db 252 MTSLSVITLLDHPQYALALRADRSRLVCGAYEBELLRYLADIAGKATATADIEVEGHLIR 311
OY 314 KGDOIILPOMLSGLDERKNACPMHVDFSROKVSHTTEGHSGLCQIHLARREIYTLKE 373
Db 312 AGESEVIIVANSIANDGTVYEDPDALDITHRSAIRHHILAGFCVHCTQGNLARLELEVITNA 371
OY 374 WLTRIPDESTA-PCAQIOHKSG-IVSVOQLPLVM 406
Db 372 LMDRPVTLRLAVPVEQLVLRGTITQGVNEPLVM 406

```

RESULT 15  
A48495  
linA1001 8-monoxygenase (EC 1.14.99.28) - *Pseudomonas incognita*  
C:Species: *Pseudomonas incognita*

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A48495  
R:Ropp, J.D.; Gunsalus, I.C.; Sligar, S.G.  
J. Bacteriol. 175, 6028-6037, 1993  
A:Title: Cloning and expression of a member of a new cytochrome P-450 family: cytochrome  
A:Reference number: A48495; MUID:93388536  
A:Accession: A48495  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-406 <ROP>  
A:Cross-references: GB:I23310; NID:g405542; PIDN:AAA2810.1; PID:g405543  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase  
E:242-377/Domain: cytochrome P450 homology <CYP>

Query Match 15.0%; Score 327; DB 1; Length 406;  
Best Local Similarity 26.3%; Pred. No. 2,1e-17;  
Matches 105; Conservative 73; Mismatches 177; Indels 44; Gaps 10.

Search completed: October 4, 2000, 12:36:44  
Job time: 24429 sec

RESULT 15  
A48495  
linA1001 8-monoxygenase (EC 1.14.99.28) - *Pseudomonas incognita*  
C:Species: *Pseudomonas incognita*



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 13:04:53 ; Search time 27.02 Seconds  
(Without alignments)  
474.815 Million cell updates/sec

Title: US-09-246-451-13

Perfect score: 2177  
Sequence: 1 TTTETIQSNANLAPLPHPVPE.....IVSGVOALPLWDPATRAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2165	99.4	414	1 CPXA_PSEPU	P00183 pseudomonas
2	591.5	27.2	414	1 Y4VG_RHISN	O53215 rhizobium s
3	376	17.3	393	1 YH85_MYCTU	O53936 mycobacteri
4	371	17.0	410	1 CPXM_BACME	O06069 bacillus me
5	359	16.5	398	1 YZ18_MYCTU	O53563 mycobacteri
6	355	16.3	396	1 YZ1B_BACSU	O34374 bacillus su
7	350.5	16.1	405	1 CPXM_BACSU	P27632 bacillus su
8	346.5	15.9	428	1 YM66_MYCTU	O50696 mycobacteri
9	340	15.6	402	1 CPXE_STRGO	P18337 streptomyc
10	334.5	15.4	410	1 CPXL_BACME	P14762 bacillus me
11	334	15.3	410	1 CPXY_BACSU	O08469 bacillus su
12	333.5	15.3	405	1 CPXA_SACER	P33271 saccharopol
13	328	15.1	405	1 CPXE_STRGO	P18336 streptomyc
14	323	14.8	428	1 CPXL_PSEPU	P33006 pseudomonas
15	318.5	14.6	405	1 YC56_MYCTU	O11062 mycobacteri
16	317.5	14.6	399	1 FAS1_RHOFA	P46373 rhodococcus
17	314	14.4	412	1 CPXH_STRGR	P53554 bacillus su
18	305.5	14.0	395	1 BIOT_BACSU	P26911 streptomyc
19	302	13.9	438	1 Y180_MYCTU	O08464 mycobacteri
20	301	13.8	402	1 NOR_FUSOX	P23295 fusarium ox
21	294.5	13.5	408	1 NOR_CYLTO	O12599 cylindrocac
22	293	13.5	381	1 CPXG_STRSO	P33266 streptomyc
23	291.5	13.4	433	1 YZ45_MYCTU	P17856 mycobacteri
24	286.5	13.2	403	1 NOR_CYLTO	O00616 cylindrocac
25	286.5	13.2	404	1 CPXJ_SACER	O00441 saccharopol
26	276.5	12.7	404	1 CPXO_SACER	P48655 saccharopol
27	273	12.5	414	1 Y778_MYCTU	P77903 mycobacteri
28	263	12.1	400	1 CPXP_RHISN	P55544 rhizobium s
29	251	11.5	401	1 CPXP_BRAJA	O59203 bradyrhizob
30	237	10.9	402	1 Y766_MYCTU	P77992 mycobacteri
31	235	10.8	396	1 YM76_MYCTU	O59571 mycobacteri
32	225	10.3	436	1 THCB_RHOER	P43492 rhodococcus
33	221.5	10.2	400	1 YV21_MYCTU	O08362 mycobacteri

34	218	10.0	422	1 CPXC_AGRTE	P24466 agrobacteri
35	216	9.9	429	1 CPXR_BRAJA	O59204 bradyrhizob
36	204.5	9.4	489	1 YM68_MYCTU	O59572 mycobacteri
37	196	9.0	405	1 CPXP_BACSU	O34926 bacillus su
38	190	8.7	467	1 CPXR_RHISN	P55543 rhizobium s
39	178	8.2	368	1 CPXM_SULSO	O55080 sulfolobus
40	163	7.5	1048	1 CPXB_BACME	P14779 bacillus me
41	155.5	7.1	503	1 CP30_SHEEP	O29496 ovis aries
42	155.5	7.1	531	1 CP27_HUMAN	O02318 h cytochrom
43	152	7.0	407	1 CPXD_AGRTE	P24467 agrobacteri
44	151	6.9	492	1 CP33_PIG	O02390 sus scrofa
45	146.5	6.7	489	1 CP33_RABIT	P00182 oryctolagus

## ALIGNMENTS

```
RESULT 1
ID CPXA_PSEPU STANDARD; PRT; 414 AA.
AC P00183:
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME P450-CAM (EC 1.14.15.1) (CAMPHOR 5-MONOOXYGENASE).
GN CAMC OR CYP101.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86111751.
RA Unger B.P., Gunsalus I.C., Slinger S.G.;
RT "Nucleotide sequence of the Pseudomonas putida cytochrome P-450cam
RT gene and its expression in Escherichia coli.";
RL J. Biol. Chem. 261:1158-1163(1986).
RN [2]
RP SEQUENCE OF 385-414 FROM N.A.
RX STRAIN=ATCC 17453.
RC MEDLINE; 90130389.
RX Koga H., Yamaguchi E., Matsunaga K., Aramaki H., Horinouchi T.;
RT "Cloning and nucleotide sequences of NADH-putidaredoxin reductase
RT gene (camA) and putidaredoxin gene (camB) involved in cytochrome
RT P-450cam hydroxylase of Pseudomonas putida.";
RL J. Biochem. 106:831-836(1989).
RN [3]
RP SEQUENCE.
RX MEDLINE; 83030788.
RA Hanlu M., Ames L.G., Yasunobu K.T., Shastri B.A., Gunsalus I.C.;
RT "Amino acid sequence of the Pseudomonas putida cytochrome P-450. II.
RT Cyanogen bromide peptides, acid cleavage peptides, and the complete
RT sequence.";
RL J. Biol. Chem. 257:12664-12671(1982).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX MEDLINE; 86143817.
RA Finzel B.C., Weber P.C., Hardman K.D., Salamee F.R.;
RT "Structure of ferricytochrome c' from Rhodospirillum rubrum at
RT 1.67-A resolution.";
RL J. Mol. Biol. 186:627-643(1985).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE; 86059514.
RA Poulos T.L., Finzel B.C., Gunsalus I.C., Wagner G.C., Kraut J.;
RT "The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450.";
RL J. Biol. Chem. 260:16122-16130(1985).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE; 98019009.
RA Schlichting I., Jung C., Schulze H.;
RT "Crystal structure of cytochrome P-450cam complexed with the (1S)-
RT camphor enantiomer.";
RL FEBS Lett. 415:253-257(1997).
```

RA [7]  
X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RX MEDLINE: 98313255.  
RA Vidakovic M., Sligar S.G., Li H., Poulos T.L.;  
RT "Understanding the role of the essential Asp251 in cytochrome p450cam  
RT using site-directed mutagenesis, crystallography, and kinetic solvent  
RT isotope effect."; Biochemistry 37:9211-9219(1998).  
RL Biochemistry 37:9211-9219(1998).  
RN [8]  
STRUCTURE BY NMR.  
RX MEDLINE: 97459726.  
RA Moura C., Bondon A., Simoneaux G., Jung C.;  
RT "1H-NMR study of diamagnetic cytochrome P450cam: assignment of heme  
RT resonances and substrate dependance of one cysteinate beta proton."; FEBS  
RT Lett. 414:203-208(1997).  
RL -1- FUNCTION: INVOLVED IN A CAMPHOR OXIDATION SYSTEM.  
CC -1- CATALYTIC ACTIVITY: CAMPHOR + PUTIDAREDOXIN + O(2) = 5-EXO-  
CC HYDROXYCAMPHOR + OXIDIZED PUTIDAREDOXIN + H(2)O.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M12546; AAA25760.1; -;  
CC EMBL: D00528; BAA00412.1; -;  
DR PIR: A25660; OAPSCP.  
DR PDB: 2CXP; 15-APR-91.  
DR PDB: 3CXP; 15-APR-91.  
DR PDB: 4CXP; 15-JUL-91.  
DR PDB: 5CXP; 15-JUL-91.  
DR PDB: 6CXP; 15-JUL-91.  
DR PDB: 7CXP; 15-JUL-91.  
DR PDB: 8CXP; 15-JUL-91.  
DR PDB: 1CP4; 15-JAN-93.  
DR PDB: 2CP4; 15-JAN-93.  
DR PDB: 3CP4; 15-JAN-93.  
DR PDB: 4CP4; 15-JAN-93.  
DR PDB: 5CP4; 16-SEP-98.  
DR PDB: 6CP4; 16-SEP-98.  
DR PDB: 1NOO; 08-MAR-96.  
DR PDB: 1PHA; 31-OCT-93.  
DR PDB: 1PHB; 31-OCT-93.  
DR PDB: 1PHC; 31-OCT-93.  
DR PDB: 1PHD; 31-OCT-93.  
DR PDB: 1PHE; 31-OCT-93.  
DR PDB: 1PHF; 31-OCT-93.  
DR PDB: 1PHG; 31-OCT-93.  
DR PDB: 1AKD; 19-NOV-97.  
DR PFAM: PF00067; P450; 1.  
DR PROSITE: PS00066; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW 3D-structure.  
FT INIT\_MET 0  
FT BINDING 357  
FT CONFLICT 55  
FT CONFLICT 276  
FT CONFLICT 361  
FT CONFLICT 407  
FT TURN 16  
FT TURN 20  
FT HELIX 22  
FT STRAND 23  
FT TURN 28  
FT TURN 29  
FT TURN 32  
FT HELIX 34  
FT HELIX 36  
FT TURN 38  
FT TURN 48  
FT STRAND 53  
FT HELIX 58

FT STRAND 62  
FT HELIX 65  
FT TURN 76  
FT TURN 78  
FT STRAND 81  
FT TURN 82  
FT HELIX 83  
FT HELIX 84  
FT TURN 88  
FT TURN 90  
FT TURN 99  
FT TURN 102  
FT TURN 105  
FT HELIX 107  
FT HELIX 108  
FT HELIX 119  
FT TURN 121  
FT TURN 145  
FT TURN 146  
FT STRAND 147  
FT HELIX 149  
FT HELIX 150  
FT TURN 153  
FT TURN 154  
FT HELIX 156  
FT HELIX 157  
FT TURN 167  
FT TURN 168  
FT HELIX 171  
FT HELIX 171  
FT HELIX 193  
FT HELIX 213  
FT TURN 219  
FT TURN 225  
FT STRAND 227  
FT TURN 228  
FT STRAND 229  
FT STRAND 231  
FT STRAND 232  
FT HELIX 235  
FT TURN 251  
FT TURN 252  
FT HELIX 253  
FT HELIX 266  
FT HELIX 276  
FT HELIX 278  
FT TURN 291  
FT STRAND 292  
FT STRAND 295  
FT STRAND 297  
FT STRAND 301  
FT TURN 308  
FT TURN 310  
FT STRAND 312  
FT TURN 314  
FT STRAND 317  
FT STRAND 322  
FT TURN 324  
FT TURN 325  
FT TURN 329  
FT TURN 331  
FT TURN 335  
FT TURN 340  
FT TURN 341  
FT TURN 349  
FT TURN 353  
FT HELIX 355  
FT TURN 358  
FT HELIX 360  
FT STRAND 377  
FT STRAND 382  
FT TURN 383  
FT STRAND 385  
FT STRAND 391  
FT STRAND 392  
FT STRAND 396  
FT STRAND 398  
FT STRAND 403  
FT HELIX 408  
FT SEQUENCE 410  
SQ 414 AA; 46538 MW; ECA610293A9D6207 CRC64;

Query Match 99.4%; Score 2165; DB 1; Length 414;  
Best Local Similarity 99.5%; Pred. No. 4.8e-156;  
Matches 412; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTTTSSNNANLAPLPVPHVEHLVDFDMDTNPSTLSAGVDEANAVLDOESNVPDLVWTRCNG 60  
DB 1 TTTTSSNNANLAPLPVPHVEHLVDFDMDTNPSTLSAGVDEANAVLDOESNVPDLVWTRCNG 60  
QY 61 GHWIATRGOLIREAYDYVHNFSSCEPFIPREAGEAYDFTPTSMDDPEORQFRALANQVVG 120  
DB 61 GHWIATRGOLIREAYDYVHNFSSCEPFIPREAGEAYDFTPTSMDDPEORQFRALANQVVG 120  
QY 121 MPVVDKLENRIQELACSLIESLRPOGQCNFTEDYAPFPFIRIFMLLAGLPEEDIPHLKYL 180  
DB 121 MPVVDKLENRIQELACSLIESLRPOGQCNFTEDYAPFPFIRIFMLLAGLPEEDIPHLKYL 180  
QY 181 TDQMTFPGSGMTFAEAKALYDYLPIITDOROKPGTDAISIVANGOVNRPITTSDEARK 240

Db	181	TDQMPRPGSMTEFAEKALDYLLPIIEQRQRKGTDAISVANGQVNGRITSDENR	240
Qy	241	MGGLLVGGDLTVNPFLSFSEMEFLAKSPENROELIERPERIPACEBELLRRPSLVADGRI	300
Db	241	MGGLLVGGDLTVNPFLSFSEMEFLAKSPENROELIERPERIPACEBELLRRPSLVADGRI	300
Qy	301	LTSDEFFGVOLKKKDDOLLPMOMSGLDERKACPMHVDPSROKXSHTFEGSHLCIGQ	360
Db	301	LTSDEFFGVOLKKKDDOLLPMOMSGLDERKACPMHVDPSROKXSHTFEGSHLCIGQ	360
Qy	361	HLARERIIVTLTKEMLTIRIPDPSIAPGAOIOHKSGIVSGVQALPLVMDPATTTKAV	414
Db	361	HLARERIIVTLTKEMLTIRIPDPSIAPGAOIOHKSGIVSGVQALPLVMDPATTTKAV	414
RESULT	2		
YAVG_RHISN	YAVG_RHISN	STANDARD:	PRT: 414 AA.
AC	053215:		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	13-DEC-1998 (Rel. 37, Last annotation update)		
DE	PROBABLE CYTOCHROME P450 YAVG (EC 1.14.14.-).		
GN	YAVG.		
OS	Rhizobium sp. (strain NGR234).		
OG	Plasmid sym PNGR234a.		
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;		
OC	Rhizobiaceae; Rhizodium.		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE: 97305956.		
RA	Freiberg C.A., Fellay R., Balroch A., Broughton W.J., Rosenthal A.,		
RA	Perret X., "Molecular basis of symbiosis between Rhizobium and legumes."		
RT	Nature 387:394-401(1997).		
RL	(2)		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE: 96389014.		
RA	Freiberg C., Perret X., Broughton W.J., Rosenthal A.,		
RA	"Sequencing the 500-kb GC-rich symbiotic replicon of Rhizobium sp.		
RT	NGR234 using dye terminators and a thermostable 'sequenase': a		
RT	beginning."		
RL	Genome Res. 6:590-600(1996).		
CC	-I- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE		
CC	MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED		
CC	COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.		
CC	-I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL: Z68203: CAA92422.1: -		
DR	EMBL: AE000101: AAB91895.1: -		
DR	HSSP: P00183: 3CP4.		
DR	PFAM: PF00067: P450_2.		
DR	PROSITE: PS00086: CYTOCHROME_P450; FALSE NEG.		
KW	Hypothetical protein: Oxidoreductase; Monooxygenase;		
KW	Electron transport; Heme; Plasmid.		
FT	BINDING 364 364 HEME (BY SIMILARITY).		
FT	SEQUENCE 414 AA: 45810 MW: 88859869315BBA4 CRG64:		

	Query Match	27.2%	Score 591.5;	DB 1;	Length 414;
	Best Local Similarity	34.1%;	Pred. No. 1.8e-37;		
	Matches 142; Conservative	65;	Indels 39;	Gaps 9;	
QY	13 PLPHVPELVDEDFMKNPSNLSA---GYQEAMAVLQESNVDDLYW----TRCNGSHWTA 65				
	:               : : :				

```

Db 13 PIPDHVPALVHNHSLSTSPOMATTPNGDHPHAACVCHDDGPRTFFSPSTRBGRGWTW 72
QY 66 TRGOLLREAVEDYRHFSSSECFPIPREAGEAVDPIPTSMDDPEORQFRALANOVVGMYPD 125
Db 73 TRARDQRRVLEDTETFFSSHRSIFASALGEHMVPIPLDLPDPAHVGFALLNPIFFSSRVL 132
QY 126 KLEKRIGELACSLIESL-ROGOCNFTEDEAEPPIPIIFMLAGLPE-----EDI 174
Db 133 ALEPTIHARACALLDCAKETSCDVKMDKALPPTFSVFLSLFGLSQRSEVLVGVWSDI 192
QY 175 PLKAYLTQDMRPDGSMTFAFAKCALVDYLIPITEO---RROKPCDASISVANGVNC 230
Db 193 LH-----GN--AEKRRAAARSVAALIDEMAAMRRSPAVDETFVQAQITEG 237
QY 231 RPIITSEAKRMFGILLVGLDITVNVPLFSFMEFLAKSPERHOLEIERPERIPACCELTAR 290
Db 238 RSLIIEEVRGICGVFLVAGADITVAALIGFDMAYIARNPKHQOLLERNPARGIAGAEELR 297
QY 291 RFSLVADGRILITSDYERHGVOLKKGDDOILPOMLSGIDERRKMACPMHVDFSROKVSHTTF 350
Db 298 AVSTVOILIRVATKQIEEGVPIREGDVSVCAMIANNDPEEFCPMTIDARODNOHTAF 357
QY 351 GHGSLGICGHLARREIIVTLKEULTPIPRESTAPG-AQLOHNSGIVSGVQALPIVW 406
Db 358 GYPHPLCHGHLARREIYIGLEWELAPIPARINEGAPLTH-GGVHFGISNIIITW 413

```

RESULT	3
YH85_MCYCTU	
ID	YH85_MCYCTU
AC	053936;
DT	15-FEB-2000 (Rel. 39, Created)
DT	15-FEB-2000 (Rel. 39, Last sequence update)
DT	15-FEB-2000 (Rel. 39, Last annotation update)
DE	PUNATIVE CYTOCHROME P450 RV1785C (EC 1.14.-.-).
GN	RV1785C OR MTW049.07C.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RP	[1]
RA	SEQUENCE FROM N.A.
RC	STRAIN=H37RV;
RX	MEDLINE: 98295987.
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA	Gordon S.V., Eigmeier K., Gae S., Barry C.E. III, Tekala F.,
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA	Taylor K., Whitehead S., Barrett B.G.;
RT	"Deciphering the biology of Mycobacterium tuberculosis from the
RT	complete genome sequence.";
RL	Nature 393:537-544(1998).
CC	-1 SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and that statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
CC	EMBL; AL022021; CAA17707.1; -
DR	TUBERCULIST; RV1785C; -
DR	PFAM; PF00067; P450; 2.
DR	PROSITE; PS00086; CYTOCHROME_P450; 1.
KW	Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
PT	BINDING 342 342 HEME (BY SIMILARITY).
SO	SEQUENCE 393 AA: 43541 MW: 86CDID508471FA8 CRC64;

	Query Match	17.33:	Score 376:	DB 1:	Length 393:
	Best Local Similarity	28.7%:	Pred. No.3e-21:		
	Matches 110:	Conservative 70:	Mismatches 183:	Indels 20:	Gaps
Qy	31 PSLNSAGVGEAMAYLOESNVDLVWTRCNGCMWTATRGOLIREAEDYRHFSSECPFIPR	90			
Db	23 PMADVGV--GMKTLIRDAG--PVVF--MNGWYL-TTREVDLALRLNPVFPSSKKALQP-	74			
Qy	91 EAGEYEYFIPTSMDDPEOROFRALANOVGMPVYDKLENRIQELACSLIESLRPOGCCNF	150			
Db	75 -PGNDLPVPVPLAFPPPEHTRYRILIOPFSPAAISKALPISLRHRTVMATIAIGRCGEA	133			
Qy	151 TEDVAEPPPIITPMILNGLPEDIPHLKYLTLDOK-----TPRDGSMTFAEKKAEXDLIP	206			
Db	134 MADLANLEPFLVLTLGLPLEDDRLDGKMDAVIAMSDRPHPEADVAARELETETLA	193			
Qy	207 IIEORROKPGTDAISIYANGOVNRPIITSDEAKRMFGILLVGGLDTVFNFTSFMELAK	266			
Db	194 MYAEERRRRPGFDVLSQV---QIGDPDISETLEVLSHLILAGIDLYTAANGVEFELLEAR	250			
Qy	267 SPEHRQELIERPERKIIPAACEBLRL-RPSLYADGRILTSDYEFHGVLKKGDQLLPOMLS	325			
Db	251 RPOLRAMLRDNPKOIRVFEIEIVRLAESAPVAPRVPTTEPVTYGGMTLPAGSPVRICAAV	310			
Qy	326 GLDERKNACPMHNDFSRQKVSHTFFGSHCLCGOHLARREIYTLKEWLTRIPDESIAIP	365			
Db	311 NRDSGDANSSTDELVDWGCVHRHMVFGGGPRHCISHLARLETLLVGMNLQIPDEFELAP	370			
Qy	386 --GAOIQHKSQIVSGVOLPLVM 406				
Db	371 DYAPEIRFPSKSPA-LKNLPLRW 392				

  

RESULT	4
CPIXM_BACME	
ID CPIXM_BACME	STANDARD; PRT; 410 AA.
AC 006063;	
DT 01-FEB-1995 (Rel. 31, Created)	
DT 01-FEB-1995 (Rel. 31, Last sequence update)	
DT 15-FEB-2000 (Rel. 39, Last annotation update)	
DE CYOCROHOME P450(MEG) (EC 1.14.99.-) (STEROID 15-BETA-HYDROXYLASE)	
DE (STERIOD 15-BETA-MONOOXYGENASE).	
GN CYP106A2.	
OS Bacillus megaterium.	
OC Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC Bacillus/Staphylococcus group; Bacillus.	
[1]	
RN SEQUENCE FROM N.A.	
RC STRAIN-ATCC 13368;	
RX MEDLINE; 94049677.	
RA Rauschenbach R., Isenrhagen M., Noeske-Jungblut C., Boldt W., Siewert G.;	
RA "Cloning sequencing and expression of the gene for cytochrome P450meg, the steroid-15 beta-monooxygenase from Bacillus megaterium ATCC 13368.",	
RL Mol. Gen. Genet. 241:170-176(1993).	
[2]	
RN CHARACTERIZATION.	
RP STRAIN-ATCC 13368;	
RC MEDLINE; 79194051.	
RX Berg A., Ingelman-Sundberg M., Gustafsson M.;	
RA "Purification and characterization of cytochrome P-450meg.",	
RT J. Biol. Chem. 254:5264-5271(1979).	
[3]	
RN CHARACTERIZATION.	
RP STRAIN-ATCC 13368;	
RC MEDLINE; 82091079.	
RA Berg A., Ratfer J.J.;	
RT "Studies on the substrate specificity and inducibility of cytochrome P-450meg.",	
RL Biochem. J. 196:781-786(1981).	
-1- FUNCTION: HAS THE CAPACITY TO HYDROXYLATE CERTAIN STEROIDS IN THE 15-BETA POSITION.	

[illegible]



RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krog A., Moles A., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skellern S., Squares S., Squires R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrett B.G.,  
RT "Deciphering the Biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AL022022; CAI17755.1; -.  
DR TUBERCULIST; RV3518C; -.  
DR PFAM: PF00067; P450.1.  
DR PRINTS: PR00359; BP450.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
FT BINDING 340 HEME (BY SIMILARITY).  
SQ SEQUENCE 398 AA; 44398 MW; BCFP3C23ECB5767F CRC64;

Query Match 16.5%; Score 359; DB 1; Length 398;  
Best Local Similarity 28.2%; Pred. No. 5; 8e-20;  
Matches 107; Conservative 69; Mismatches 169; Indels 34; Gaps 8;

QY 55 WTECN-----GGHWATGSGOLIREAYEDYRHFS-----ECPTFRAGENVYDPT 101  
DB 26 WMRANQVDFDRNGLAASYQAVIDERQPELFSNAGIRPDQALP-----WMT 76  
QY 102 SMPPEORFALANOVGAPVVDKLENRIOELACSLIESLRPOGOCNFTEDYAEPPPIR 161  
DB 77 DMDPAILLRKRLVNGFTTKRKVKDKASTALCDTLIDAVCERGEDPFDALAPLMA 156  
QY 162 IFMLLAGL-PEEDIPHLK-----YTDQMTPRDGSMTFAEAKALYDYLPIIEQRQ 213  
DB 137 VIDMDLGVREQDMFLRMSDDLVTFLSHVSODEFOITM-DAFAAYNDFTRAITARRA 195  
QY 214 KPGTDAISIVANGVNGRPTISDEAKRMFGLLVYGGLDVVNFSPFMEKLASPEAROE 273  
DB 196 DPTDDLVSIVSVSVDGERISDDELVMETLLILIGDETRHTLSGGTEQLLRKDDQWL 255  
QY 274 LIERPERIPACEELLRRFSLVAD-GRILTSDEYFHVQLKKGDOILLPQMLSGIDERN 332  
DB 256 LQNDPSLLPAIEMLRWTPAVKNCMVLTADTEFHGTALCAGEKMLLEPSANFDEAVF 315  
QY 333 ACPMHVDENQKVSHTTEGSHLCLGQHLARREIYTLKEWLTTRIPDFSIAPGAQI--Q 390  
DB 316 CEPEKFPVQNRPNSHLAFGCTHCLGNQALRELISLATERVLRRLPDLRLVADSVPL 375  
QY 391 HKSGIVSGVQALPLYMDPA 409  
DB 376 RPAFVSGLESMPVFTPS 394

RESULT 6  
YJTB\_BACSU STANDARD; PRT; 396 AA.  
AC 034374;  
DT 15-FEB-2000 (Rel. 39, Created)  
DT 15-FEB-2000 (Rel. 39, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PUTATIVE CYTOCHROME P450 YJTB (EC 1.14.-.-).  
GN YJTB  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.

RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Rivoita C., Soldo B., Lazarevic V., Joris B., Manuel C., Karamata D.;  
RT "A 35.7 kb DNA fragment from Bacillus subtilis chromosome containing a  
RT putative 12.3 kb operon involved in hexonolate catabolism and a  
RT perfect catabolite-responsive element."  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF015825; AAC46317.1; -.  
DR EMBL: Z99110; CAB13078.1; -.  
DR SUBTILIST: BG13195; YJTB.  
DR PFAM: PF00067; P450.1.  
DR PRINTS: PR00359; BP450.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
FT BINDING 349 HEME (BY SIMILARITY).  
SQ SEQUENCE 396 AA; 44990 MW; 9A89CFL12613DBCFB CRC64;

Query Match 16.3%; Score 355; DB 1; Length 396;  
Best Local Similarity 30.1%; Pred. No. 1; 1e-19;  
Matches 99; Conservative 54; Mismatches 158; Indels 18; Gaps 7;

QY 71 IREAYEDYRHFSCEPTFRAGENVYDPTSMPPROFALANOVGAPVVDKLENR 130  
DB 56 VKKVVGDKELFSS-C-MPOQTSSIGNSI-INMDPPKHTKIRSVNKAFTPRVAKQMEPR 111  
QY 131 IOELACSLIESLRPOGOCNFTEDYAEPPPIRIFMLLAGLPEEDIPHLKYLTDQMT 185  
DB 112 IOETDELIOFQGRSEFDLVHDSPLPVIVISELGLVPSAMHEORFAMSDLLVSTPKD 171  
QY 186 -RPDGSMTFAE---AKEALYDYLPIIEQRQKPGTDAISIVANGVNGRPTISDEAKR 240  
DB 172 KSEAEKAPLFEERKCEELAAFAFAGLIEKRNNRPEGDIIISLVEAETGKLSGEPLP 231  
QY 241 MFGLLVGLDITYVNFISFSEMEFLAKSPENRQELIERPERIPACEELLR-RESLVADGR 299  
DB 232 FCTLLVAGNETTNNLISNAYSILETPGYEEELRSHPELMPQAVEBALRFRAPAVLRR 291  
QY 300 ILNSDYFHVQVLKKGQOILLPQMLSGIDERNKACPMHVDENQKVSHTTEGSHLCLG 359  
DB 292 IAKRDTLIGHLIKEGMVLAFAVANRDEKKEFDRPHMFDIRRPHNPHIAGHHCGLG 351  
QY 360 QHLARREIYTLKEWLTTRIPDF---SIAP 385  
DB 352 APLARLEANIATLSLSAFPMMECVSITP 380

RESULT 7  
CPXM\_BACSU STANDARD; PRT; 405 AA.  
AC P27632;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE CYTOCHROME P450 109 (EC 1.14.-.-) (ORF405).  
GN CYP109.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M23;



[illegible]

```

OY 346 SHTTGHSHTLGGHRLARREITVLKEMLTIPPEFSIA-PEAOIOHSGI-VSGVOLP 403
    | | | | | | | | | | : : : | | | | | : : : | | | | |
DB 340 RHVAFGSHVQGLPKARVELQIAETLRLPDLRLAVPHEETPFPGDMAITGVHSLP 399
OY 404 LVW 406
    : |
DB 400 IAW 402

RESULT 10
CPXL_BACME STANDARD: PRT: 410 AA.
AC PI4762:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450(BM-1) (EC 1.14.14.1).
GN CYP106.
OS Bacillus megaterium.
OC Bacteria: Firmicutes; Bacillus/Clostridium group;
    Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 1-25.
RC STRAIN-ATCC 14581;
RX MEDLINE: 90089408.
RA He J.S., Ruettinger R.T., Liu H.-M., Fulco A.J.:
RT "Molecular cloning, coding nucleotides and the deduced amino acid
    sequence of P-450BM-1 from Bacillus megaterium.";
    Biochim. Biophys. Acta 1009:301-303(1989).
RN [2]
RP SEQUENCE OF 1-25 FROM N.A.
RX MEDLINE: 95355495.
RA He J.S., Liang Q., Fulco A.J.:
RT "The molecular cloning and characterization of BM1p1 and BM1p2
    proteins, putative positive transcription factors involved in
    barbiturate-mediated induction of the genes encoding cytochrome
    P450BM-1 of Bacillus megaterium.";
    J. Biol. Chem. 270:18615-18625(1995).
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
    MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
    COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation --
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch)
CC
DR EMBL: X16610; CAA34612.1; -
DR EMBL: S79230; AAC60495.1; -
DR PIR: S07764; O4BS6M.
DR HSSP: P33006; 1CPT.
DR PFAM: PF00067; p450. 1.
DR PRINTS: PR00359; BP450.
DR PROSITE: PS00086; CYTOCHROME_P450. 1.
DR Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
KW BINDING 356 HEME.
FT SEQUENCE 410 AA; 47460 MW; C9AE293E76745387 CRC64;
SQ

Query Match 15.4%; Score 334.5; DB 1; Length 410;
Best Local Similarity 27.7%; Pred. No. 4.2e-18;
Matches 96; Conservative 57; Mismatches 170; Indels 23; Gaps 7;

OY 63 WIATGOLIREVEYEDRIHSSSE---CPIPRAGAGAYOFIP---TSDMPROGFRALA 115
    | : : : | | | | | | | | | | | | | | | | | | | | | |
DB 45 MNVPEYEHAKVQLSNIDPFSSDGOQTTTFVGDNSKKKSTSPYTNLTNLPDPHRAKSL 104

OY 116 NOVGMVAVDKLENIQELACLISLRPOGOCNTFEDYAEPPFIRIFMLAGLPEEDIP 175

```

Db 105 AAAPPSRLKNMEPRKQIADLVEALQKNSTINIVDDSSPPPSVIADLFQVYKDRY 164  
 QY 176 HLKYLTDQMTRPDGSMTFAEAKE-----ALYDYLPIIEORROKPGTATISVANGOV 228  
 Db 165 QFKRWVDILFQRPDQERLEIEDEKORAGAEYQYLYPIYIEKRSNLSDDILISDILQAEV 224  
 QY 229 NGRPTISDEAKRMGGLLVGGLDTVNF-L-SFSMEFLAKSPREHROELIERPERIPAACEE 287  
 Db 225 DGEFTFDEEIVHAATMLLGACVETTSIAINMEFSLYDOKSLYSLELRNRELAPKAVEE 284  
 QY 288 LLR-RPSLVADGRLTSDYEFHGVOALKKGDOILLPQMLGLDERKKNACPMHVDPSR-QKV 345  
 Db 285 MLRYREHISRRDRTVAKODNELLGVKKLGDVVIAMMSACNMDETMEPEPSVDIHRPTNK 344  
 QY 346 SHRTFGHSLCIGOHILAREIIVTLKEMLTR-----IDFSIAP 385  
 Db 345 KHLTFGGRPHRCIGAPLARLEMKILLEAFLEAFSHTEPFEDFELEP 390

RESULT 11  
 CPXY\_BACSU STANDARD; PRT: 410 AA.  
 AC 008469:  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE CYTOCHROME P450 (EC 1.14.-.-).  
 GN CYP4 OR CYP107J1.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / BGSC1A1;  
 RX MEDLINE: 97431495.  
 RA Belitsky B.R., Gustafsson M.C.U., Sonenshein A.L., von Wachenfeldt C.;  
 RT "An lrp-like gene of Bacillus subtilis involved in branched-chain  
 RT amino acid transport.";  
 RL J. Bacteriol. 179:5448-5457(1997).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE: 97453479.  
 RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,  
 RA Duesterhoeft A., Ehrlich S.D.;  
 RT "Sequence of the Bacillus subtilis genome region in the vicinity of  
 RT the lrp operon reveals two new extracytoplasmic function RNA  
 RT polymerase sigma factors Sigv and Sigz.";  
 RL Microbiology 143:2939-2943(1997).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Y11043; CAA11937.1; -  
 DR EMBL: 093876; AAB80898.1; -  
 DR EMBL: 299117; CAB14615.1; -  
 DR HSSP: Q00441; IOXA.  
 DR SUBTLIST: BG11929; CYP4.  
 DR PRAM: PR00067; P450.1.  
 DR PRINTS: PR00359; BP450.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 DR Oxidoreductase; Monooxygenase; Heme.  
 FT BINDING 359 359 HEME (BY SIMILARITY).  
 SQ SEQUENCE 410 AA; 47384 MW; 03598858A4761AB CRC64;

Query Match 15.3%; Score 334; DB 1; Length 410;

Best Local Similarity 27.2%; Pred. No. 4, 6e-18;  
 Matches 86; Conservative 69; Mismatches 139; Indels 22; Gaps 5;  
 QY 104 DPDEQOFALANOVGMVYVDLENRIQELACLSLELRPOQOCFTDYAPPRIRLF 163  
 Db 100 DDPDHNRLRTLVOKAFTHTRIITDLEDKIOTIADSLDKVQPKMKMLVDDYAFRLPIYI 159  
 QY 164 MLAGLPEEDIPHLKYLTDQMTRPDGSMTFAEAKEL-----YDYLPIIEQRQ 213  
 Db 160 SEMGILPELDROKFRVWSQAI-----IDFSAPRELRQENDHLGFEVLESLVKKRR 213  
 QY 214 KPGTDAISIVANGOVNGRPTISDEAKRMGGLLVGGLDTVNF-L-SFSMEFLAKSPREHROE 273  
 Db 214 EPAGDILSALIQASEGTOISTEELYSMILILVAGHETVTNLTMTFYALMCNHQDLER 273  
 QY 274 LIERPERIPAACELLRRSLV--ADGRILTSDYEFHGVOALKKGDOILLPQMLGLDERK 331  
 Db 274 LRQOPDLMSALIEEARLEHSPVPELTTRTWAEPFLIHGOEIKRKDYIILISLASANDERK 333  
 QY 332 NACPMHVEDSRQKVSHTTGHGSHLCIGOHILAREIIVTLKEMLTRIPPSIAPGAQIOH 391  
 Db 334 FRPADLPDIERKKNRHLIACGHGHPQLGAGQANLAEKIAISTLLRRCPIQL-KGEKKOM 392  
 QY 392 K---SGIVSGVALPL 404  
 Db 393 KWKGNFLMRALBELPL 408

RESULT 12  
 CPXY\_SACER STANDARD; PRT: 405 AA.  
 AC P33271;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE CYTOCHROME P450 107B1 (EC 1.14.-.-) (P450CVIIB1).  
 GN CYP107B1.  
 OS Saccharopolyspora erythraea (Streptomyces erythraeus).  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;  
 OC Saccharopolyspora.  
 RN (1)  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 172-203 AND 302-321.  
 RC STRAIN=NRRL 2338;  
 RX MEDLINE: 92121109.  
 RA Andersen J.F., Hutchinson C.R.;  
 RT "Characterization of Saccharopolyspora erythraea cytochrome P-450  
 RT genes and enzymes, including 6-deoxyerythronolide B hydroxylase.";  
 RL J. Bacteriol. 174:725-735(1992).  
 CC -1- FUNCTION: NOT KNOWN, PROBABLY INVOLVED IN THE CATABOLISM OF  
 CC OCTANE AND GUAIACOL. IT DISPLAYS A WEAK ACTIVITY IN THE O-  
 CC DEALKYLATION OF 7-ETHOXYCOUMARIN.  
 CC -----  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M83110; AAA26483.1; -  
 DR PIR: BA2606; B42606.  
 DR HSSP: Q00441; IOXA.  
 DR PFAM: PF00067; P450.1.  
 DR PRINTS: PR00359; BP450.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 DR Oxidoreductase; Monooxygenase; Electon transport; Membrane; Heme.  
 FT BINDING 352 352 HEME (BY SIMILARITY).  
 SQ SEQUENCE 405 AA; 45238 MW; 71C93CEC1PDC53FD CRC64;





SQ SEQUENCE 405 AA: 44580 MW: 72DEAEC6B688FA48 CRC64;

Query Match 14.6%; Score 318.5; DB 1; Length 405;

Best Local Similarity 29.5%; Pred. No. 6,7e-17; Matches 102; Conservative 51; Mismatches 160; Indels 33; Gaps 9;

```
QY 62 HMIATRGQLREAYEDYRHFSSCEPIPREAGE----AYDFIPTSN-DPEQORFRALA 115
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 46 YYLSRHADYWSAARDHQFESS-AQGLTVNYGLEMIGLHDTPPMWQDPVHTFERKLY 104
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 116 NOYVGMPPYVDKLENRIQELACSLIESLRPOGOCNFTEDYAEPPPIRIEMLLAGLPEEDIP 175
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 105 SRGFTPROVETVEPTVRKFVVERLEKLRANGGDDIVTELFKPLPSMVVAHYLGVPPEEDWT 164
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 176 HLKYLTDQWRP---DGSMTFA-EAKEALDYLLPIIEORROKPGTDATIS--IVANGQVN 229
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 165 QPDGWTQAIYANANAVDGAATTGALDAVGSMMAYFTGLIERRRTPEADDAISHLVNAGVGAD 224
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 230 GRPITSDEAKRMFGLLVGGLDVTVNFLSFSMEFLAKSPEROELIERPERIPACEEEL 289
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 225 GDTAGTLSILAFTFTWVTGNDFTVGMGSMFLHRRPDQRLLDDPREGIPDAVEELL 284
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 290 RRFSLYAD-GRILTSYEFHGVOLKKGDOILLPQMLSGLDERK-----NACPMHV 338
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 285 RLTSPTVOGLARTTRTDVTTIGDFTIPAGRRLVLLYGSANRDEROYGPDAELDVTRCPRNI 344
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 339 DFSROKVSHTTFGHSGLICQHLARREIIVTLKEWLTIRIPDFSIA 384
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 345 -----LTFSHGAHCLGAAAAARQCRAVLTLLARCPDFEVA 381
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

Search completed: October 4, 2000, 13:04:55  
Job time: 1686 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:39:37 ; Search time 79.83 Seconds  
(without alignments)  
359,568 Million cell updates/sec

Title: US-09-246-451-13

Perfect score: 2177  
Sequence: 1 TTERIQSNANLAPLPPIHVE.....IVSGVQALPLVWPATTKAV 414

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_12:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organeller:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	364	16.7	396	2 052544	052544 amycolatops
2	363.5	16.7	416	2 087605	087605 streptomyc
3	362	16.6	397	2 059523	059523 micromonosp
4	345	15.8	410	2 09XSP9	09XSP9 streptomyc
5	344.5	15.8	404	2 054302	054302 streptomyc
6	344.5	15.8	407	2 059819	059819 streptomyc
7	337	15.5	411	2 032460	032460 actinomadur
8	327	15.0	406	2 059723	059723 pseudomonas
9	325.5	15.0	376	2 031785	031785 bacillus su
10	323.5	14.9	420	2 09ZHQ1	09ZHQ1 streptomyc
11	321	14.7	407	2 09XSP8	09XSP8 streptomyc
12	315	14.5	406	2 087192	087192 streptomyc
13	314	14.4	400	2 09XSP7	09XSP7 streptomyc
14	313	14.4	400	2 09XDB0	09XDB0 mycobacteri
15	311.5	14.3	410	2 059831	059831 streptomyc
16	307	14.1	310	2 032927	032927 mycobacteri
17	304.5	14.0	357	2 024727	024727 nocardioide
18	303	13.9	417	2 059910	059910 streptomyc
19	299	13.7	388	2 P77977	P77977 streptomyc

20	299	13.7	406	2 085697	085697 streptomyc
21	291	13.4	351	2 052572	052572 amycolatops
22	283.5	13.0	337	2 052816	052816 amycolatops
23	280.5	12.9	406	2 087675	087675 amycolatops
24	280	12.9	421	2 052561	052561 amycolatops
25	272	12.5	403	2 052560	052560 amycolatops
26	269.5	12.4	386	2 059921	059921 streptomyc
27	266.5	12.2	310	2 P96562	P96562 amycolatops
28	263.5	12.1	437	2 09ZPC0	09ZPC0 mycobacteri
29	262	12.0	395	2 09XST2	09XST2 streptomyc
30	256	11.8	398	2 087674	087674 amycolatops
31	254	11.7	426	2 09ZFC3	09ZFC3 mycobacteri
32	251.5	11.6	411	2 09X8Q3	09X8Q3 streptomyc
33	250	11.5	411	2 060005	060005 s putative
34	249.5	11.5	511	2 052569	052569 amycolatops
35	244	11.2	406	2 052823	052823 amycolatops
36	240	11.0	433	2 09XCC6	09XCC6 streptomyc
37	225	10.3	398	2 059079	059079 amycolata a
38	224	10.3	391	2 052822	052822 amycolatops
39	222.5	10.2	391	2 087673	087673 amycolatops
40	204.5	9.4	419	2 050242	050242 agrobacteri
41	204	9.4	397	2 052802	052802 amycolatops
42	182.5	8.4	119	2 085655	085655 streptomyc
43	179	8.2	120	2 085653	085653 streptocallo
44	172.5	7.9	313	2 09X4I8	09X4I8 myxococcus
45	164	7.5	120	2 085650	085650 streptomyc

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	396 AA.
052544				
ID	052544			
AC	052544			
DT	01-JUN-1998 (TREMBLrel. 06, Created)			
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)			
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)			
DE	CYTOTOXIC P450 MONOOXYGENASE.			
OS	Amycolatopsis mediterranei.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinobacteriales; Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-5699.			
RA	KIM C.G., YU T.W., FRYHLE C., HANDA S., FLOSS H.G.;			
RL	J. Biol. Chem. 267:10041-10048, 1992.			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-5699.			
RA	AUGUST P.R., TANG L., YOON Y.J., NING S., MUELLER R., YU T.W.,			
RA	TAYLOR M., HOFFMANN D., KIM C.G., ZHANG X., HUTCHINSON C.R.,			
RA	FLOSS H.G.;			
RL	Chem. Biol. 5:0-0(0002).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-5699.			
RA	AUGUST P.R., TANG L., YOON Y.J., NING S., MUELLER R., HUTCHINSON C.R.,			
RA	FLOSS H.G.;			
RL	Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF040570; AAC01709.1; -			
DR	HSSP: 000441; 10XA.			
DR	PRAM: PR00067; P450. 1.			
DR	PRINTS: PR00359; BP450.			
KW	Monooxygenase.			
SQ	SEQUENCE 396 AA: 44293 MW: BA875129 CRC32:			

Query Match 16.7%; Score 364; DB 2; Length 396;  
Best Local Similarity 29.08; Prod. No. 1.9e-21;  
Matches 117; Conservative 59; Mismatches 183; Indels 44; Gaps 14;  
QY 27 DWYPSNLASGVQENAVAVLQESNVPLVWTRCNGH--WIATRGQLIREAYEDYRHSSE 84

```
Db 15 DKEDPAVEDSLKEE-RPLAKMVPD-----GHGWMVSSYSELVREVLSDLR-FSHS 64
OY 85 CP---FIPREAGYADYPIPT-----SMDPEQROFALANOVYGMVVDKLEMRIQ 132
Db 65 CEVGHFPPVTHQGV---IPRHPIPGMFTIMDPPEHTRRYKLLTGTEVTVARASRLIPRAE 121
OY 133 ELACSLIESLRPG-OCNFTEDYAEPPPIRIFMLAGLPEED---IPHLKYLTDOMTRP 187
Db 122 AVAAEIEVWARAKAPADVMDAFKPLVLEMLGELVGLPYEERDRVPAYVATLLHDAEADP 181
OY 188 DSGMTFAEAKELXYDLPIIEORROKPGTDAISIVANGVNRPTSDAKMFGLLY 247
Db 182 AEA---AAAEVAGKPFDEYIERRORPDDLLSSLVT-----EDLTQELRNITVTLTF 233
OY 248 GGLDTVNFISFSEMEFLAKSPENHOELIERPERIPACEELLRRFSL--VADGRILTSDY 305
Db 234 AGETTEGALATGVFALHHTDQALRAPEPKLDAIEELLKLYVNOYHTTRTALEDV 293
OY 306 EPHVOLKKGDOILLPOMLSGLDERKNACPMHVDFSROKVSHTTFGSHLCLGHLARR 365
Db 294 KLGELIKKGDYVYSIPAA NRDPAKFCPAFLDIEDTSGHVAFGHIGCLGONLARI 353
OY 366 EIIVTLKEMLTIRIPDSIA-PGAQIOHK-SGLVSGVALPLW 406
Db 354 ETRAGFTALLRAPELRLAVPADEVPLRLKGSVSKLPLVSM 396

RESULT 2
087605 PRELIMINARY: PRT: 416 AA.
AC 087605;
DT 01-NOV-1998 (TREMblrel, 08, Created)
DT 01-NOV-1998 (TREMblrel, 08, Last sequence update)
DT 01-NOV-1999 (TREMblrel, 12, Last annotation update)
DE CYTOCHROME P450 MONOOXYGENASE.
GN PICK OR PIKC.
OS Streptomyces violaceus (Streptomyces venezuelae), and
OC Streptomyces venezuelae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC15439;
RA BETLACH M.C., KEALEY J.T., BETLACH M.C., ASHLEY G.W., MCDANIEL R.;
RT "Characterization of the macroliide P450 hydroxylase from Streptomyces
RT venezuelae which converts narbomycin to picromycin.";
RT Biochemistry 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-S. VENEZUELAE; STRAIN=ATCC15439;
RA MEDLINE: 99051447.
RA XUE Y., WILSON D., ZHAO L., LIU H.-W., SHERMAN D.H.;
RT "Hydroxylation of macroliactones YC-17 and narbomycin is mediated by
RT the PIK-encoded cytochrome P450 in Streptomyces venezuelae.";
RT Chem. Biol. 5:661-669(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-S. VENEZUELAE; STRAIN=ATCC15439;
RA MEDLINE: 98445333.
RA XUE Y., ZHAO L., LIU H.-W., SHERMAN D.H.;
RT "A gene cluster for macroliide antibiotic biosynthesis in Streptomyces
RT venezuelae: architecture of metabolic diversity.";
RT Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
DR EMBL: AF087022; AAC64105.1; -.
DR EMBL: AF079139; AAC68886.1; -.
DR HSSP: Q00441; 10XA.
DR PFAM: PF00067; P450.1.
DR PRINTS: PR00359; BP450.
KW Monooxygenase.
SQ SEQUENCE 416 AA; 46038 MW; CAD6EBA0 CRC32;
```

```
Query Match 16.7%; Score 363.5; DB 2; Length 416;
Best Local Similarity 28.3%; Pred. No. 2.3e-21;
Matches 97; Conservative 62; Mismatches 163; Indels 21; Gaps 7;

OY 76 EDYRHSSECPPIPRRAGEYDPIPTSMDDPEQROFALANOVYGMVVDKLEMRIQEL 135
Db 72 KDMR--NSTTPTLEAERAAHNMLES--DPPRTRLRKLVAREFTIRRELLRPROVEIV 127
OY 136 CSLSIEL--RPGOCNFTEDYAEPPPIRIFMLAGLPEEDIPHLKYLTDOMTRPDGSMF 193
Db 128 DGLVDMALAAPDORADLMESLAMPRLITVISELGVPEPDRAPFRWTDFAVFPDDPA 187
OY 194 AEAKELXYDLPIIEORROKPGTDAIS-IVANGVNRPTSDAKRMFGLLVGLDPT 252
Db 188 QTAMMSGYLSRLIDSKRGODEDLISALVRTSDRGSRLTSELLGMAHILLVAGHET 247
OY 253 VVNFISFSEMEFLAKSPENHOELIERPERIPACEELLRRFSLVADRILTSDEF----- 307
Db 248 TVNLIANGVALLSHPDOLAALRADMTLLDGAVEMLR-----YEGPVRSATYRFPVEPV 302
OY 308 --HGVOLKKGDOILLPOMLSGLDERKNACPMHVDFSROKVSHTTFGSHLCLGHLARR 365
Db 303 DLDGYIIPADYVLYVLADAHRTPERPPRPHRDITRDTAGHLAFGHGIFCIGAPLARL 362
OY 366 EIIVTLKEMLTIRIPDS--IAPGAQIOHKSGLVSGVALPLW 406
Db 363 EARIAVALLERCPLDALDVSPGELVWYPMIRGLKALPIRW 405

RESULT 3
059523 PRELIMINARY: PRT: 397 AA.
ID 059523;
AC 059523;
DT 01-NOV-1996 (TREMblrel, 01, Created)
DT 01-NOV-1996 (TREMblrel, 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel, 12, Last annotation update)
DE CYTOCHROME P450-LIKE PROTEIN (EC 1.14.14.1).
GN MYCG.
OS Micromonospora griseorubida.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micromonosporinae; Micromonosporaceae;
OC Micromonospora.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A11725;
RA MEDLINE: 95107242.
RA INOYE M., TAKADA Y., MUTO N., HORINOCHI S., BEPPU T.;
RT "Characterization and expression of a P-450-like mycinamicin
RT biosynthesis gene using a novel Micromonospora-Escherichia coli
RT shuttle cosmid vector.";
RT Mol. Gen. Genet. 245:456-464(1994).
DR EMBL: D16098; BAA03672.1; -.
DR HSSP: Q00441; 10XA.
DR PFAM: PF00067; P450.1.
DR PRINTS: PR00359; BP450.
SQ SEQUENCE 397 AA; 44331 MW; DID592A6 CRC32;
```

```
Query Match 16.6%; Score 362; DB 2; Length 397;
Best Local Similarity 28.5%; Pred. No. 2.8e-21;
Matches 105; Conservative 57; Mismatches 168; Indels 38; Gaps 8;

OY 63 WIATRQQLRBEAYEDYRHHSSCPFI-----PRAGEAYDPIPTSMDDPEQROFRA 113
Db 44 WLVTN-----YEDVRAVLGGRFVGRGSMTRDEPRTREPENKGGLSMDPEHSRLRR 96
OY 114 LANOVYGMVVDKLEMRIQELACSLIESLRPOG-CNFTEDYAEPPPIRIFMLAGLPEE 172
Db 97 LVYKATFARASLSIRRAEINHELVDQMAATGQPADLVAMARQLPVAVITCELLGVESA 156
OY 173 DIPHLKYLTDOMTRPDGSM-----TFAEAKELXYDLPIIEORROKPGTDAISIV 223
Db 157 D-----HDFRTFMGAPLSTAETVAEEMGEAAEQAVAYMGDLIDRRRKEPTDLVSL 209
```

OY	224	ANGQVNGRPITSDAKRRKRGLLLYGDLTVYNLFSTFSEMEFLASKSPHKROLIERPERIPA	263
Dd	210	VQARQODSDLSBOELLDLGAICLVAGYEETTTQADPVLMTLRPELRDLDRLPELIPS	269
OY	284	ACEELLRFRESL---VADGRILTSDYERFHGVQKKGDQDILLPQMUSGLDERKNACPMHVF	340
Dd	270	AVEELTRFWPLPGVGTAFFRVAVEDYTTLRGVTIRAGEEYLVASTGCANDQAFDPADRIDV	329
OY	341	SROKVSHTTFGHSGHLCLGQHARRELTIVTLKEMLTRIIPDFSIA-PCGAQIQHKSQ-.IYSC	398
Dd	330	DRTPNQHGIFGHGVHNHCAGPARLAVELQVALVELLQSLRPGIRLGIIPTQLRWSEGMILRG	389
OY	399	VQALPLVM 406   :	
Dd	390	PLLELPVVW 397   :	
<hr/>			
RESULT	4		
O9X5P9			
ID	O9X5P9	PRELIMINARY;	PRT; 410 AA.
AC	O9X5P9;		
DT	01-NOV-1999 (TREMBLrel_12, Created)		
DT	01-NOV-1999 (TREMBLrel_12, Last sequence update)		
DT	01-NOV-1999 (TREMBLrel_12, Last annotation update)		
DE	CYTROCHROME P450 HYDROXYLASE ORF3.		
OS	Streptomyces lavendulae.		
OC	Bacteria; Firmicutes; Actinobacterlia; Actinobacteridae;		
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-NREL 2564;		
RX	MEDLINE; 99201491.		
RA	MAO Y.Q., VAROGU M., SHERMAN D.H.;		
RT	"Molecular characterization and analysis of the biosynthetic gene		
RT	cluster for the antitumor antibiotic mitomycin C from Streptomyces		
RL	lavendulae NREL 2564."		
Chem.	Biol. 6:251-263(1995).		
DB	EMBL; AF127374; AAD28449.1; -		
SO	SEQUENCE 410 AA; 45887 MW; 507C7F38 CRC32;		

Query Match	15.8%	Score 345;	DB 2;	Length 410;
Best Local Similarity	27.0%;	Fred. No. 6.8e-20;		
Matches 113;	Conservative 69;	Mismatches 193;	Indels 44;	Gaps 14;
Oy	9	ANLAPLP-P-HVEHILVFEDMYNPNSL	SAGVGEAAVVLQESNVDPDLVYTRCGNGHMIATR	67
Db	15	AGEAAVYFFPHADRL--EPDPY-----	WEPLRREPLDQVILPYGGEMLATR	60
Oy	68	GGLIREAVEDYKHFESSE---CPFIREAGEAVDFTP-----	TSMDPREGQRPALAN	116
Db	61	YQDVAVRAVAD--RRFSROLAVAPGABR-----	FLPHQPPRAVLSVEGDPHARLRILVG	112
Oy	117	QVYVGNPVYDKLENRIQELACLSTIESLRQG--QCNTEDYAEFFPIRIFMLLAGLREEDIP		175
Db	113	KVFPRRVEDMPLLIQTRADGLDLMEMGPRADIVEDPFSFAVSMICELLAGVPERKK		172
Oy	176	NKLYLTDMQTRPDGSMTPFAEAKE---ALVDYLPIITEORQKRPQVDAISIVANGQVNGRP		232
Db	173	RCVMSDALILTTT-AHTPAQVRDYMKNQHDILGGLVAGVACRVAPRTADLIGSLVTANDEEDK		231
Oy	233	ITSDAKRRMFGLLVGLGDLTVVNFISFSMEFLAKSPENHOBELIERPERIPRACSELLRRF		292
Db	232	LTEGELVRLAEILLAGVETSAQIPNFIYULFRHPOLLERIRNDHDLIPDAVEELRLPV		291
Oy	293	SL-VADG--RLITSYEFHGVOLTKGDDOILPQMSLGDERRKNACPMHNDFSROKVSHT		349
Db	292	PCLGYVDGPFRTATEDVELGGLVYRAGETVYPSMGANDRPDLFTTPRDEDLARRRPNHIG		351
Oy	350	FGHSHLCLQGLHARREIIVTLKEMLTIRPDFSIA--PGAQIQHKSGI--VSGVQALPLVW		406
Db	352	FGAGHGHICGADLAVELQDILTTLTFFRRPRRLRYLVRPEESLSMKSGLWRCGNHTRPVW		410

RESULT	5			
054302		PRELIMINARY;	PRT;	404 AA.
AC	054302			
AC	054302;			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)			
DE	CYTOCHROME P450.			
GN	RAPN.			
OS	Streptomyces hygroscopicus.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
CC	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NRRL 5491;			
RX	MEDLINE; 95372374.			
RA	SCHWECKE T., APARICIO J.F., MOLNAR I., KOENIG A., KHAW L.E.,			
RA	HAYDOCK S.F., OLIVNYK M., CAFFERY P., CORTES J., LESTER J.B.,			
RA	BOEHM G.A., STAUNTON J., LEADLAY P.F.;			
RT	"The biosynthetic gene cluster for the polyketide immunosuppressant			
RT	rapamycin.";			
RL	rapamycin.;"			
RL	proc. Natl. Acad. Sci. U.S.A. 92:7839-7843(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NRRL 5491;			
RA	MOLNAR I., APARICIO J.F., HAYDOCK S.F., EE KHAW L., SCHWECKE T.,			
RA	KOENIG A., STAUNTON J., LEADLAY P.F., STAUNTON J., LEADLAY P.F.;			
RL	Gene 0:0-0(0).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NRRL 5491;			
RA	APARICIO J.F., MOLNAR I., SCHWECKE T., KOENIG A., HAYDOCK S.F.,			
RA	EE KHAW L., STAUNTON J., LEADLAY P., LESTER J.B., BOEHM G.A.,			
RA	STAUNTON J., LEADLAY P.F.;			
RL	Gene 0:0-0(0).			
DR	EMBL; X86780; CAA60465.1; -			
DR	HSSP; 000441; 10XA.			
DR	PFAM; PF00067; P450; 1.			
DR	PRINTS; PR00359; BP450.			
SO	SEQUENCE 404 AA; 45071 MW; 05AB94DF CRC32;			

	Query Match	15.8%	Score 344.5;	DB 2;	Length 404;	
	Best Local Similarity	28.6%;	Pred. No.7.3e-20;			
	Matches 102; Conservative	63;	Mismatches 177;	Indels 15;	Gaps	9;
QY	63 WIATGGLIREVEYEYRIHNSSECFP----	IIPRAGEAYDF-IITSMDPRQGOFALANOV	118			
	:   :	: : :   :   :   :	:	:	:   :	
Db	50 WLAASMEDAKFVDPDR-FSRSATIGKDVPRKPAIQQDVPYIMLDDPREHTLRRAVTGA		108			
QY	119 VGMPPVDKLNNYIQELACSLIESLPPOG-QCNFTEDYAPFPPIRFIMLAGLPEEDIPL	177				
	:   :	: : :   :   :   :	:	:	:   :	
Db	109 LTRSMEARPTGTGEVADLLIDMKLAKGARPMADMEDIAFLPEITIMICELLGVPIEQTFE	168				
QY	178 KYLTDOMTRPDCSGM---FAEKALKLYDLIPTIDQRORPGTDAISIVANGCVNCRPT	234				
	:   :	: : :   :   :   :	:	:	:   :	
Db	169 RFWMSOML-SNGAYSQEVMMAAGOSLSYLTLSELIAERRKODTNDDLGSIVRARDDDDRS	227				
QY	235 SDEARFMGGILLVGSDIRPVNFLESMEFLASKSPEIROBLIERPETRAPACEILRRSTL	294				
	:   :	: : :   :   :   :	:	:	:   :	
Db	228 ETELGFVAUTLLTAGETGANAIINSVYTLLTHPELBALARLDLSLIIPAVALDELRLTIPI	287				
QY	295 VADG---RLTSNDYEFHGVLKKGOQLLHPQLSGIDEKKNAKPMMHVDFSROKVSHTTGP	351				
	:   :	: : :   :   :   :	:	:	:   :	
Db	288 AKQAQSVNRBAVEDVELSGTVKAGEVVAAIOQTSAINTDPRVYUYPHEIDIHRHSINHHMSG	347				
QY	352 HCSHLCLGHOLARRREITYLKEMLTRIPDEFSTA-PGAOTQHKS--TVSGVALPLYW	406				
	:   :	: : :   :   :   :	:	:	:   :	
Db	348 HAHHCMGNQAVRMQMOTALGSLISRIPALRPAREVEDEPKLFJRGVLVSLEMLPLTW	404				

RESULT	6
ID	Q59819
AC	Q59819;
	PRELIMINARY; PRT; 407 AA.
DT	01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE	CYTCHROME P450 (EC 1.14.14.1).
GN	OLEP.
OS	Streptomyces antibioticus.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces,
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE: 9525619.
RA	RODRIGUEZ A.M., OLANO C., MENDEZ C., HUTCHINSON C.R., SALAS J.A.;
RT	"A cytochrome P450-like gene possibly involved in Oleandomycin
RT	biosynthesis by Streptomyces antibioticus.";
RL	FEMS Microbiol. Lett. 127:117-120(1995).
CC	-I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR	EMBL: L37200; AAA92553.1; -.
DR	HSSP: Q0041; IOXA.
DR	PROSITE: PS00067; CYTOCHROME_P450; 1.
DR	PFAW: PF00067; P450; 1.
DR	PRINTS: PR00359; BP450.
KW	Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme,
FT	BINDING 356
FT	SEQUENCE 407 AA; 44957 MW; 04411c60 CRC32; HEME (BY SIMILARITY).

	Query Match	15.8%;	Score 344.5;	DB 2;	Length 407;	
	Best Local Similarity	28.1%;	Pred. No. 7.4e-20;			
	Matches 104;	Conservative 63;	Mismatches 168;	Indels 35;	Gaps 11;	
Oy	60 GGHWTATRGQLIREAYEDYRIHFSEC---PFIPIREAGEAYDFIPT-----SMDPPQ	108				
	I : I :	I :	I :	I :	I :	I :
Dd	50 GTAVLVTRNSDARIYVLGDSR-FSTAADDPATR-----KFPITRPEDGVIAQDPDH	101				
	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
Oy	109 ROFRALANOVGMYPVDKLLENRIQELIACLSLELRPOG-CNFTEYAEFFPIRTFMILA	167				
	: : : : : :	: : : : : :	: : : : : :	: : : : : :	: : : : : :	: : : : : :
Dd	102 TRLRLTVCKAKFAARVVEEMRPVRSLVPSLIDDMVAHNGSPADVFLPAVPFVAVICELL	161				
	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
Oy	168 GLPEEDIHLKLYLTQM---PRDDSMPTAEKKEALYDLI---PIIROROKPGTDALS	221				
	: : : : : :	: : : : : :	: : : : : :	: : : : : :	: : : : : :	: : : : : :
Dd	162 GPLLEDRLDRFESDAMSTR---LTAELTORVOODMYMVGDIAGVRAQRAPFDLLG	217				
	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
Oy	222 IVANGOVNGRPTSDSEAKRMFGILVGLDYVNFLSFSEMEFLAKSPENROETIERPERI	281				
	: : : : : :	: : : : : :	: : : : : :	: : : : : :	: : : : : :	: : : : : :
Dd	218 ALALATDNDHDHTKGELIVNMGVSLLIAGHETSVMQITNLVHLLTERKKYESLVADPALV	277				
	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
Oy	282 PACCELLARRSLVADG---RLTSDYEFGVOLKGGDILLPOMISGLDERKNACPMNV	338				
	: : : : : :	: : : : : :	: : : : : :	: : : : : :	: : : : : :	: : : : : :
Dd	278 PAAVEEMERYPLVSAGSFVRVATEDVELSTVTBAGECVCVVHFGASANDDEVEFHADEL	337				
	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
Oy	339 DFSRKVSHHTTGSGHSCLCOHLARRETIYLKMLPIIPFSLA-PGAOLQHSG-TV	396				
	: : : : : :	: : : : : :	: : : : : :	: : : : : :	: : : : : :	: : : : : :
Dd	338 DFHRNRNHIAFGHANHCIGAOLRLEIQELALSALVRFRPFLDLAEPVAGLKWKOGMLI	397				
	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
Oy	397 SGVQALPLWV	406				
	: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	: : : : :
Dd	398 RGLERQIVSW	407				
	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
RESULT	7					
ID	032460					
PRELIMINARY:		PRT:	411 AA.			
AC	032460:					
DT	01-JAN-1998 (TREMBLrel. 05, Created)					
DT	01-JAN-1998 (TREMBLrel. 05, last sequence update)					
DT	01-NOV-1999 (TREMBLrel. 12, last annotation update)					
DE	ORF 10.					
OS	Actinomodura hibisca.					
OC	Bacteria; Firmicutes; Actinobacterlia; Actinobacteridae;					

OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;  
OC Actinomadura.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PI57-2;  
RX MEDLINE: 97480928.  
RA DAIRI T., HAMANO Y., IGARASHI Y., FURUMAI T., OKI T.;  
RT "Cloning and nucleotide sequence of the putative polyketide synthase  
genes for predimethyl biosynthesis from Actinomadura hibiscus.";  
RL Biosci. Biotechnol. Biochem. 61:1445-1453(1997).  
DR HSSL: D87924; BAA23153.1; ~.  
DR HSSP: O00411; 10XA.  
DR PFAM: PF00067; p450.1.  
DR PRINTS: PR00359; BP450.  
SO SEQUENCE 411 AA; 44860 MW; 98A628B6 CRC32;

Query Match	15.5%	Score 337	DB 2	Length 411
Best Local Similarity	27.0%	Fred. No. 3e-19		
Matches 117	Conservative	60	Mismatches 196	Indels 60
				Gaps 17
QY	4	TIOSANLAPRPHPHPEHVEFEDMYNP----	SNLSAGVOEAMAVLOESNVDLWTRCN	59
Db	9	TVDPREDVTPAPFRPRD-----	DFGQPCHEARLRASDPVAKVLP-----	T 50
QY	60	GGH-WIATRCGLIRPATEDYRHFSSEC---	PIPRPAGCAFDYFTS-----	MDPE 107
Db	51	GDHAWVTRYADYRVTSD-RRFSEKAVTRPGAPR-----	LIPMOROSKSLVIMDPPE	102
QY	108	QROFALNOVAVGMVPEVVKLENRIEELACSLIESLRPG-	OCNPFEDYAEPPIPIFMIL	166
Db	103	HTRKAKIVSRFAFARVARGCMRAHVADLTSGFEDVENEHCPADLAIHLALPIPVVICM	162	
QY	167	AGLPEDIPHLKYLTDQMTRPDSMTFEAKR-----	ALYDYLPIIQROROKPTDAI	220
Db	163	LGVPEDPRFRODWTDRML-TIGAPALAOADEIKAAVGRKRYLAEILDAKTAAPADILL	221	
QY	221	SIVANGQVNGRPITSDEAKRMGCL-LVYGLDTVYNPLSFSMEFLAKSPENQELIERPE	279	
Db	222	SLSNAHADG--LSEELLTFGMTLLAAGYHTTAALHTSYHLLREPSRYARREDPS	279	
QY	280	RIPACEELLRRFSLVADG---RLTSDYEFHGVOLAKGDOIILPOMLSGLDERKKNACP	335	
Db	280	GIPAVEELL-RYCGIGGAGAIRIAVEDVEGTVLRAGEAVIPLFNANRDPEVADP	338	
QY	336	MHVDSRQKVSHTTGGHSHLCLGCHLARRELIYTLKEMLTPIPFSTA-PSAQQHKSG	394	
Db	339	BEILDGRDNPHIALGHGHCYCLGAPLARLEIQVLETVLVERTPALRLAIDADITWRG	398	
QY	395	IV-SGVQALPLWV	406	
Db	399	LAFARPDALPIAM	411	
RESULT	8			
ID	059723	PRELIMINARY:	PRT:	406 AA.
AC	Q59723:			
DT	01-NOV-1996 (TREMblrel. 01, Created)			
DT	01-NOV-1996 (TREMblrel. 01, last sequence update)			
DE	01-NOV-1999 (TREMblrel. 12, last annotation update)			
DR	CYTCHROME P450 LIN (EC 1.14.14.1).			
GN	LINC.			
OS	Pseudomonas incognita.			
OC	Bacteria; Proteobacteria.			
NC	[1]			
RP	SEQUENCE FROM N.A.			
RA	ROPP J.D., GUNSAIUS I.C., SLIGAR S.C.;			
RL	Submitted (JAN-1994) to the EMBL/Genbank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.			
DR	EMBL: L23310; AAA25810.1; -			
DR	HSSP: P33006; ICP.			
DR	PROSITE: PS00086; CYTOCHROME P450:1.			





Db 324 NRDSAKFDPDELDPRPTLRPHMTFGMHGHHGLCAPLATMELEVASSTLLTTRPALRLDV 383  
QY 384 APGAQIQHKSGIVGVALPLVW 406  
Db 384 PREDVSNMTSITMRYPLALPVTM 406

RESULT 13  
Q9X9P7 PRELIMINARY; PRT; 410 AA.  
AC Q9X9P7;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, last sequence update)  
DE 01-NOV-1999 (TReMBLrel. 12, last annotation update)  
DE NIKE PROTEIN.  
GN NIKE  
OS Streptomyces tendae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TRUE901;  
RA BRUNTNER C., LAUER B., SCHWARZ W., MOEHRL E., BORMANN C.;  
RT "Molecular characterization of co-transcribed genes from Streptomyces  
RT tendae Tu901 involved in the biosynthesis of the peptidyl moiety of  
RT the peptidyl nucleoside antibiotic nikkomycin.";  
RL Mol. Gen. Genet. 0:0-0(0).  
DR EMBL: Y18574; CAB46536.1; -  
SQ SEQUENCE 410 AA; 45884 MW; 9B52CB74 CRC32;

Query Match 14.4%; Score 314; DB 2; Length 410;  
Best Local Similarity 27.1%; Pred. No. 2.1e-17;  
Matches 98; Conservative 64; Mismatches 180; Indels 20; Gaps 9;

QY 63 WIATRGQLIREAYEDYRHFSSSEC-PRIP-----REAGEAYDFIP-----SMDPEQR 109  
Db 47 WLVLKHLARKLLADPRVSADRLHAPAPRGRLTAEOPTATERVRRLSTRSMIHLDGDEHG 106  
QY 110 QFRALANOVGVGVVVKLENRIOELA-CSLIESLRPOGOCNFTEDYAEPPRIEMLAG 168  
Db 107 AHKKILTGFESLRRIALRRVQVEIDVRSIDEMLAPOPADIVEHNSQAVPSLVCCLG 166  
QY 169 LPREDIPHL-KYLLDQMTPRDGSMTFAEAKELYDYLPII-BOOROKPGDAISIYANG 226  
Db 167 VPHEDQRDFEMAGMLVSRKSYSTRERAAADALNDFLEDLYTEKEKEPTDDLIGRLIAR 226  
QY 227 QVNGRPITSDAKRMFGLLVGGIDTVVNFLSFSMEFLAKSPENROELIERPERIPACE 286  
Db 227 NRRTPVMTHEIYGTAVMLLIAGHOTANNISLGVALLENPEKKARIADPSLPPAIE 286  
QY 287 ELLRRSLV--ADGRILTSYEFHGVOLKKGDDILLPQMLSGDERKNACPMHVESRQK 344  
Db 287 EMLRYSVVENAPARVATEDIEIGVATIRKDEGIIVSGLAADMDEVEFHRDLDFERGA 346  
QY 345 VSHTTGSHGSHLGLGHLARREIIVTLKEMLTIRPDSIA-PCAQIQHKSGI-VSGVAL 402  
Db 347 RHNVAGSYGVHCGGLGMALHARLEIYFETLLRRVPGSLAVPAEELRYKDDAGIYGYRV 406  
QY 403 PL 404  
Db 407 PV 408

RESULT 14  
Q9XDBO PRELIMINARY; PRT; 400 AA.  
AC Q9XDBO;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, last sequence update)  
DE 01-NOV-1999 (TReMBLrel. 12, last annotation update)  
DE CYTOCHROME P450.  
GN PIPA.

OS Mycobacterium smegmatis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;  
OC Actinomycetales; Corynebacteriinae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC2155;  
RX MEDLINE; 99287823.  
RA FOUJIN P., DUCROCQ V., HALLIER-SOULIER S., TRUFFAUT N.;  
RT "Cloning and characterization of the genes encoding a cytochrome P450  
RT regulatory protein (PipR) in Mycobacterium smegmatis mc2155.";  
RL J. Bacteriol. 181:3419-3426(1999).  
DR EMBL: AF102510; AAD28344.1; -  
SQ SEQUENCE 400 AA; 44747 MW; C10DD01A CRC32;

Query Match 14.4%; Score 313; DB 2; Length 400;  
Best Local Similarity 26.8%; Pred. No. 2.4e-17;  
Matches 115; Conservative 57; Mismatches 171; Indels 86; Gaps 14;

QY 26 FDMYNP--SNLSGVQDA-----W-----AVIQESNPDDV-----WTRCNG 60  
Db 12 FDTDPAFSTISDEVHEAREKSWATTPYGLAVLRYSQVNRLLKHPKLRGSSAMPANH 71  
QY 61 -----GHWATRGQLIREAYEDYRHFSSSECPIFRPAGAYDPIPTSMDPPEOROPRAL 114  
Db 72 VTGEPFAEMFAS-----WLNKEGE-----EHHRLRL 99  
QY 115 ANOVGVNPNVVKLENRIOELACSLIESLRPOGOCNFTEDYAEPPRIEMLAGLPREDI 174  
Db 100 MNPAFSPKLIGSLVPRFOALANELLIDNFAEPDRCFSEFAEPYAAVIAIMGLPREE- 158  
QY 175 PLKYLTDQ-----MTRPDGSMFAEAKELYDYLPIIEQRQKGTDAISVAN 225  
Db 159 --WKVISTESATIGLAVLRLEDLPRIEAAVORLYEYSEDLADRANPRDPMITLVN 216  
QY 226 GOVNGRP-----ITSDEAKRMFGLLVGGIDTVVNFLSFSMEFLAKSPENROELIERPERI 281  
Db 217 A-----SRDDGRLSKELRDALLILIFGFPDTRQQLAMQTKPKHNDQRLGGERDILG 273  
QY 282 PACEELLRRFSIVA-DGRILTSYEFHGVOLKKGDDILLPQMLSGDERKNACPMHVP 340  
Db 274 GKAAEEVYRVNPNYRWMTREVDEFEYEGVTLKAGTVVHLHSESAGTDP--VEPQFDI 331  
QY 341 SROKVSHTTGHSHLGLGHLARREIIVTLKEMLTIRPDSIAPGAQIQHKSGIVSGVQ 400  
Db 332 TAERKPHFGFGGVHCLGHFVARSMSALPLLARLRDPHLEPGATWLPDSG-NTGPN 390  
QY 401 ALPLVMDPA 409  
Db 391 TLPDGTTPA 399

RESULT 15  
Q59831 PRELIMINARY; PRT; 410 AA.  
AC Q59831;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, last sequence update)  
DE 01-NOV-1999 (TReMBLrel. 12, last annotation update)  
DE CYTOCHROME P450 SCA-2 (EC 1.14.14.1).  
OS Streptomyces carophilus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SANK 62585;  
RX MEDLINE; 96001248.  
RA WATANABE I., NARA F., SERIZAWA N.;  
RT "Cloning, characterization and expression of the gene encoding  
RT cytochrome P-450sca-2 from Streptomyces carophilus involved in  
RT production of pravastatin, a specific HMG-CoA reductase inhibitor.";  
RL Gene 163:81-85(1995).





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 06:41:30 ; Search time 2798.97 seconds  
(Without alignments)  
591.079 Million cell updates/sec

Title: US-09-246-451-16  
Perfect score: 927  
Sequence: 1 atgcagtaacctacatc.....gagtgtaacagcaactct 927

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pl2:\*  
11: gb\_pl3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: em\_fun:\*  
17: em\_hum1:\*  
18: em\_hum2:\*  
19: em\_in:\*  
20: em\_cm:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_sy:\*  
29: em\_un:\*  
30: em\_vl:\*  
31: gb\_hlg1:\*  
32: gb\_hlg2:\*  
33: gb\_hlg3:\*  
34: gb\_hlg4:\*  
35: em\_ba1:\*  
36: em\_ba2:\*  
37: em\_hum3:\*  
38: em\_hum4:\*  
39: gb\_pl4:\*  
40: gb\_hlg3:\*  
41: gb\_hlg4:\*  
42: gb\_hlg5:\*  
43: gb\_hlg6:\*

44: gb\_hlg7:\*  
45: em\_hlg1:\*  
46: em\_hlg2:\*  
47: em\_hlg3:\*  
48: em\_hum5:\*  
49: gb\_pl3:\*  
50: gb\_pl5:\*  
51: gb\_hlg8:\*  
52: gb\_hlg9:\*  
53: gb\_hlg10:\*  
54: gb\_hlg11:\*  
55: gb\_hlg12:\*  
56: gb\_hlg13:\*  
57: gb\_hlg14:\*  
58: gb\_hlg15:\*  
59: gb\_hlg16:\*  
60: gb\_hlg17:\*  
61: gb\_hlg18:\*  
62: em\_hlg4:\*  
63: em\_hlg5:\*  
64: em\_hlg6:\*  
65: em\_hlg7:\*  
66: em\_hum6:\*  
67: gb\_hlg18:\*  
68: gb\_hlg19:\*  
69: gb\_hlg20:\*  
70: gb\_hlg21:\*  
71: gb\_hlg22:\*  
72: gb\_hlg23:\*  
73: gb\_hlg24:\*  
74: gb\_hlg25:\*  
75: gb\_hlg26:\*  
76: gb\_hlg27:\*  
77: gb\_hlg28:\*  
78: gb\_hlg29:\*  
79: gb\_hlg30:\*  
80: gb\_hlg31:\*  
81: gb\_vil1:\*  
82: gb\_vil2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	925.4	99.8	955	5 A00740	A00740 A.rusticiana
2	925.4	99.8	955	5 A00741	A00741 A.rusticiana
3	925.4	99.8	955	14 SYNHRPER	J05552 Synthetic h
4	843.8	91.0	1072	5 A30294	A30294 Synthetic I
5	843.8	91.0	1072	5 AR000440	AR000440 Sequence
6	843.8	91.0	1072	5 AR068580	AR068580 Sequence
7	564.6	60.9	927	5 A06250	A06250 Synthetic H
8	524	56.5	924	5 E01651	E01651 CDNA encodi
9	493.6	53.2	1236	7 ATPRXCB	X71794 A.thaliana
10	490.4	52.9	1229	7 HRAHRPCC	M60729 A.rusticiana
11	481.4	51.9	1195	7 ATPEROX5	X98777 A.thaliana
12	481.4	51.9	1195	7 ATPRXR3GE	X98315 A.thaliana
13	304.4	32.8	1102	7 PTPXP2PER	X97349 P.trichocar
14	301.8	32.6	1143	7 POPP01	D83224 Populus nig
15	297.2	32.1	1371	7 PTPXP4PER	X97351 P.trichocar
16	296.4	32.0	1461	7 PTPXP1PER	X97348 P.trichocar
17	295.4	31.9	1222	7 PTPXP3PER	X97350 P.trichocar
18	292	31.5	1115	7 POPPA	D30652 Populus kit
19	290.4	31.3	1268	8 LINEPROX	L07554 Linum usita
20	275.4	29.7	1170	7 POPPB	D30653 Populus kit
21	257.2	27.7	2912	7 HRAHRPCA	M37156 A.rusticiana
22	250.4	27.0	1176	8 AF149277	AF149277 Phaseolus
23	244.4	26.4	3289	7 HRAHRPCB	M37157 A.rusticiana
24	237.6	25.6	85109	8 ATT2J13	AL132967 Arabidops

25	237.6	25.6	129757	8	ATF2K15	AL132956 Arabidops
26	237.2	25.5	3919	7	ATHL33036	AJ133036 Arabidops
27	236.4	25.6	3436	7	ATHPRXA	M58380 Arabidopsis
28	217	23.4	1169	8	ALFPXDC	L36157 Medicago sa
29	216.8	23.4	1220	7	MSRNAPFLA	X90692 M. sativa mr
30	216	23.3	1212	7	ATPPTPA2A	X99992 A. thaliana
31	215.2	23.2	1287	8	AF007211	AF007211 Glycine m
32	207	22.3	3095	7	HRAPRXC2	D90115 Hordeidish
33	204	22.0	1311	8	SSNPEROXIB	L36111 Stylosanthu
34	201.4	21.7	2459	8	AF149280	AF149280 Phaseolus
35	199.2	21.5	1238	7	MSRNAPFLB	X90693 M. sativa mr
36	195	21.0	1287	7	MSRNAPFLC	X90694 M. sativa mr
37	192.6	20.8	104386	8	AT132A17	AL161813 Arabidops
38	192.6	20.8	172675	41	AC007635	AC007635 Arabidops
39	192.6	20.8	199362	8	ATCHRIV24	AL161512 Arabidops
40	189.4	20.4	104386	8	AT132A17	AL161813 Arabidops
41	189.4	20.4	172675	41	AC007635	AC007635 Arabidops
42	189.4	20.4	199362	8	ATCHRIV24	AL161512 Arabidops
43	182.8	19.7	1467	8	PAB250121	AJ250121 Picea abi
44	180	19.4	1320	49	AF244923	AF244923 Spinacia
45	178.4	19.2	1271	8	ALFPXDA	L36156 Medicago sa

## ALIGNMENTS

RESULT 1  
A00740 955 bp DNA PAT 28-JAN-1993  
LOCUS A.rusticana synthetic gene for peroxidase.  
DEFINITION A00740.1 GI:14602  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

horsesradish.  
Armoriacia rusticana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Caprales; Brassicaceae; Armoriacia.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
SYNTHETIC GENE  
Patent: WO 8903424-A 2 20-APR-1989;  
British Bio-Technology Ltd

FEATURES  
source  
1. 955  
/organism="Armoriacia rusticana"  
/db\_xref="taxon:3704"  
11. 940  
/note="synthetic gene"

CDS  
/codon\_start=1  
/product="Horse radish peroxidase"  
/protein\_id="CAA00083.1"  
/db\_xref="GI:14603"

/translation="MOLPTFYDNCSPNPNVIRDTIVNELSDPRIASIRLHFHD  
CFVNGCDASILLDNSTFRTKDAFGNANASARGPVIDRMKAVERSCRTVSCADLT  
T1AAOOSVTLAGPSWRVPLGRDLSLOATLDLANALPAPFTLPDLKDSFNVGLNR  
SSDIVALSGHFGKNOCRFIMDRILYFNFGTGLPDLTNTYLTOTLRGICPLNGNLISA  
LVDFDLRPTTFIDNKRYVNLFEOKGLIOSDOELFSSPNATDITPLVRSRANSTQTFEN  
AFVEAMDRKGNITPLTGTOGRINCRVNSNS"

BASE COUNT 246 a 269 c 221 g 219 t

ORIGIN

Query Match 99.8%; Score 925.4; DB 5; Length 955;  
Best Local Similarity 99.9%; Pred. No. 3.3e-270;  
Matches 926; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atcagatgaaccctacatclacagcaatagctgtcccaagctgtcccaatcgttgcg 60  
|||||  
Db 11 ATCGAGTTAACCCCTACATTCCTACGACAAATAGCTGTGCCAAGCTGTCCACATCGTTCCG 70  
|||||  
QY 61 gacacaatcgtcacaagagctcagatccgatcccaagagctcgtctcaatatactcgtc 120  
|||||

Db 71 GACAAATCGTCAACGAGCTCAGATCCGATCCAGATCCGCTGCTTCATATTAGCTCTG 130  
QY 121 cacttcacatgaacgctcgtgaatggtgtgcagcgtacgatattactggaacacacc 180  
Db 131 CACTTCATACCTCCTTCGTAATGTTGGCAGCTAGCATATTACTGGCAACACCACC 190  
QY 181 agttccgacactgaagaagatgcatlccggaaacgctaacagcgccagggcttccagt 240  
Db 191 AGTTCCGCACTGAAGAGATGATTCGGAGACGCTAAACGCCCAAGGGCTTTCCAGTG 250  
QY 241 atgatcgtacatgaagcgtcgtcgttgaatcagatgccccagaacagtcaattgcaag 300  
Db 251 ATGATCGCATGAAAGGCTGCGTTGATCGCATGCCACAGCAAGTCAATGTGAGAC 310  
QY 301 ctctgacatgaactgtgcgaacagagcgtgtacactcttgcaagggagaccgttctggaagt 360  
Db 311 CTGCTGACTATTAGCTGCGCAACAGAGCGTACTTTGACAGCCGAGCCGCTCTGAGAGTG 370  
QY 361 ccgctcgttcgaagctgtacatccctcacaagcattcctagatctggccaagcgaacttgct 420  
Db 371 CCGCTCGTCGACGTGACTCCTCCTACAGAGCATTCCTAGATCTGGCCAAAGCCACTTGCC 430  
QY 421 gctccatcttcaaccctcgtcccaagcgtgaagatagcttgaagaacgttggctgaatcgc 480  
Db 431 GCTCCATCTTTCACCCCTGCCAGCTGAAGATAGCTTTGAAACGTCGCTGAATCGC 490  
QY 481 tccagtagcctgtgtcgtcgttcgcgaagacacacatttggaagaacagtgtaagtc 540  
Db 491 TCGAGTACCTTGTGCTCTGCTCCGAGAGACACACATTGTGGAAGAACAGTGAAGTTC 550  
QY 541 atcagtagagctctacaaattcagaacaactgggttaccctgaccccaagcgtgaacact 600  
Db 551 ATCATGATAGGCTTACAAATTTCAGAACACATGGGTATTACCTGACCCCAAGCTGAACACT 610  
QY 601 acgtatctcagaacactgtgaagcgttgcgtcccaactgaatggaactcagtgtaactagt 660  
Db 611 ACGTATCTCCAGACACTGAGAGAGGCTGTGCCACTGATGAGCACTCACTGACACTAGTG 670  
QY 661 gacttgatctgcggaccccaacacatcttcgataaacaagtaactgtgaatcagaaggag 720  
Db 671 GACTTTGATGTGGGAGACCCCAACCATCTTGATTAACAAGTACTATGTGAATCTAGAGGAG 730  
QY 721 cagaagaagcctgaatcagaagtgatcaagaactgtttagcagtlccagaagccactgacac 780  
Db 731 CAGAAAGGCTGTATACAGAGTGTAAAGAACTGTTAGAGTCCAAAGCCACTGACACAC 790  
QY 781 atcccaactgttgaagagtttgccttaacttaactcaactcaaaccttcttaagcctctgga 840  
Db 791 ATCCCACTGCTGAGAGTTTGTCTACTACTACATCAAACTCTTTTAAAGCCCTTCGTGGA 850  
QY 841 gccatgaccgctatggttaacatctacccctcgtgaagcggttaacgaagccagatctgtc 900  
Db 851 GCCATGACCGCTATGGGTAACTATCCCTCTGTACCGGTACCCAGAGCCAGATTCGCTG 910  
QY 901 aactcagagtggttcaacagcaactct 927  
Db 911 AACTGCAAGAGTGTCAACGACACTCT 937  
RESULT 2  
A00741/c  
LOCUS A00741 955 bp DNA PAT 02-FEB-1993  
DEFINITION A.rusticana synthetic gene (reverse complement) for peroxidase.  
ACCESSION A00741  
VERSION A00741.1 GI:14604  
KEYWORDS  
SOURCE  
ORGANISM  
horsesradish.  
Armoriacia rusticana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Caprales; Brassicaceae; Armoriacia.

REFERENCE  
AUTHORS  
Edwards, R.

TITLE	SYNTHETIC GENE									
JOURNAL	Patent: WO 8903424-A 3 20-APR-1989; British Bio-Technology Ltd									
COMMENT	#A00741 is the reverse complement of #A00740.									
FEATURES	Location/Qualifiers 1..955 /organism="Armoracia rusticana" /db_xref="taxon:3704"									
BASE COUNT	219	a	221	c	269	g	246	t		
ORIGIN										
Query Match	99.8% Score 925.4; DB 5; Length 955;									
Best Local Similarity	99.9% Pred. No. 3.3e-270;									
Matches 926; Conservative	0; Mismatches 1; Indels 0; Gaps 0;									
Oy	1	atgcagtaaacccctacac	ttctagacaaatagctctcccaagctgtccaaacgtctgc	60						
Db	945	atgcagttAACCCCTACATCT	TACGACAAATAGCTCTCCAAAGTCTCCAACTGCTGC	886						
Oy	61	gacacaatcgttcaaaagagctcagaa	tcacagatcccaagatcgcgtcttcaatatctgctc	120						
Db	885	GACACAATCGTCAAGAGACTCGAT	TCGATCCCAAGATCGCTTCATATATTAGTGTG	826						
Oy	121	caactccatgactgctctcgtcgaa	tggctgtgcagcgctagcatatctctggaacaaccacc	180						
Db	825	CACCTTCATGTAGCTCTTGTAAT	GGTTGCGCAGCTAGCATATTCTGAGCAACACACC	766						
Oy	181	agttccgacacgaaaagatgatgc	ttcgggaacgttaacagcgccagagggtcttccagtg	240						
Db	765	AGTTTCCGACATGAAAAGATCAT	TCGGGAAGCGTAACAGCGCCACGAGGGCTTTCCAGTG	706						
Oy	241	atcgaatcgcatgaagagctgcgcgt	tttagtcaagcatgtcccaagacaagctcagttgtgcagac	300						
Db	705	ATCGATTCGATGAAGGCTGCGG	TCGTAGTCAAGATGCCCAAGAACGTCACTGTGTGCAAGC	646						
Oy	301	ctgtctgactatagctctgcgcaca	agagcgttgactcttgcagcgcgacgcgtctcgtgaagtg	360						
Db	645	CTGCTGACTATATGTGCTGCGC	ACACAGAGCGCTGTGACTTTGACGCGCGACCCGCTCGAGAGTG	586						
Oy	361	ccgctccgttcgaacgtgactccct	taaggcatcttcagatctctgcgaacgcgaacttgct	420						
Db	585	CCGCTCGGTCGACGTGACTCC	CTACAGGAGATTCTTAGATCTGGCCAAACCCCAACTTGCT	526						
Oy	421	gtcccatctcttaacctgtccacc	agcttgaagataagctttaaagaacgttggtctgaatcgc	480						
Db	525	GCTCATTTCTTACCCCTGAC	CCGCCACGTGAAGATACCTTTAGAAAGTGSGGTCTGAATCGC	466						
Oy	481	tcgagtgaaactgtggtctctgtc	gcggaggaacacacatcttgaaagaaccagtgtagtctc	540						
Db	465	TCGAGTGAACCTTGTGGCTCT	GTCTGTCGCGAGGACACACATTGTGGAAGAACCAAGTGTAGGTTTC	406						
Oy	541	atcagtgatgaagctcttaacaat	ttcagcaaaccttgggttgtaacctggaacccaacgtctgaacact	600						
Db	405	ATCATGGATGAGGCTCTTACA	AATTTACGACAAACATGTGGTTTACCTGACCCCAACGCTGAACACT	346						
Oy	601	acgatctcccaagaacactgaag	agcgtctgtgcgcaacgtgaatgggaacactgaagctactagtg	660						
Db	345	ACGTATCTCCAGACACTGAG	AGGCTTGTGCCCACTGGAATGGCAACCTCACTGTCACCTAGTG	286						
Oy	661	gaacttgatctctgcgaacccca	accaactctctcgtataaacaagtactatgttgatctagaagag	720						
Db	285	GACTTTGATCTTCGGAGCC	CAACCATCTTCGATTAACAAGTACTATGTGATCTAGAGGAG	226						
Oy	721	cagaaaagcctgatatcaaga	tgatcaagaacgtgttagcagttccagaagccactgaacacc	780						
Db	225	CAGAAAGCCCTGATCTAC	AGAGTGATGAAGAACTTTAGACGATGCCAAACGCCCACTGACAC	166						
Oy	781	atcccacatggatgaagaagt	tttgcttaaccttaacttaaacacttctttaaagcctctgtgga	840						
Db	165	ATCCCACTGATGATGAAG	ATTTTCTCTACTACTACTCAAACTTTTAAACGCTTTCGTGGAA	106						
Oy	841	gccatcgagccgtatggtga	atactaacctctcgaaagcgtgttaccagaagccagaatctgctctg	900						

DB	LOCUS	SYN	27-Apr-1993
Db	105	GCACATGACCGCATGCGGTAACCTTACCGGAGGAGTACCAAGCCACAGATTTCGCTG	46
0y	901	aactgcagatggtcaacagcaactc 927	
Db	45	AACATGACAGTGTCTACACAGCAACTCT 19	
RESULT	3		
SYN	HRP	955 bp	DNA
DEFINITION	Synthetic horseradish peroxidase isoenzyme C (HRP-C) subunit		
ACCESSION	alpha-1 DNA, complete cds.		
VERSION	J05552.1	GI:208493	
KEYWORDS	peroxidase.		
SOURCE	Synthetic DNA.		
ORGANISM	Synthetic construct		
REFERENCE	artificial sequence.		
AUTHORS	1 (bases 1 to 955)		
	Smith, A.T., Santama, N., Dacey, S., Edwards, M., Bray, R.C.,		
	Thornely, R.N.F. and Burke, J.F.		
	Expression of a synthetic gene for horseradish peroxidase C in		
	Escherichia coli and folding and activation of the recombinant		
	enzyme with Ca2+ and heme		
	J. Biol. Chem. 265, 13335-13343 (1990)		
	90330683		
JOURNAL	Draft entry and computer-readable sequence for [J. Biol. Chem.		
MEDLINE	(1990) In press] kindly submitted		
COMMENT	by A.T. Smith, 13-JUN-1990.		
FEATURES	Location/Qualifiers		
source	1..955		
	/organism="synthetic construct"		
	/db_xref="taxon:32630"		
	11..940		
CDS	/note="Synthetic horseradish peroxidase isoenzyme C		
	(HRP-C) subunit alpha-1 (E.C. 1.11.1.7)"		
	/codon_start=1		
	/transl_table=1		
	/protein_id="AA07222.1"		
	/db_xref="GI:208494"		
	/translation="MQLPTPTDYNDSRPNVYDITVNELSRDPIASILRLHPD		
	CEVNGDASILLDDWTSFRTKEKDAFGNANSRGPVIDRKAAYESACPTVSCADLL		
	TIAAQSVTLAGGDSWVRPLGRDLSQAFDLIDANLPAFFTLPLKDSFRNGLNR		
	SSDIYVSGHTEFGKNOCRFLMDRLYNSTGPDPLNTYLOTLRGLDPLNGLSA		
	LVDDLRPTREIDNKYVNLLEOKGIOSOELESSPNADTIPLVASFANSYQTFPN		
	AFVAMBRMGITPLTGTGQIRLNCRVYVNS"		
BASE COUNT	246 a 269 c 221 g 219 t		
ORIGIN	1 bp upstream of HindIII site.		
Query Match	99.8% Score 925.4; DB 14; Length 955;		
Best Local Similarity	99.9%; Pred. No. 3.3e-270;		
Matches 926; Conservative	0; Mismatches 1; Indels 0; Gaps 0;		
0y	1	atgcagtaaccctcacctctcagaacaataagctgtcccaacggtgtccacatcgttcgc 60	
Db	11	ATGCGATTACCCCTACATTCCTACGACATAAGCTTCCCAACGTCTCCAACTTCGTTGC 70	
0y	61	gacacaatcgtcaacagagctcagatccgatcccaagatcgtctgtcctaataatacgtctg 120	
Db	71	GACACAATCGTCAACAGAGCTCAGATCCGATCCAGAGATCGCTTCAATATTACGCTCG 130	
0y	121	caattccatgaatcgtctgtgtaagtgttcgagagcttaacataatactctggaacaacacc 180	
Db	131	CACCTTCATGATGCTGCTGTGATAGTGGTTCGCGACGCTACCATATTACTGGACAACACCCAC 190	
0y	181	agttccgaacgaaagatgatcattcgggaaagcttaacagcgccagagggtctcccaatg 240	
Db	191	AGTTCCGACACTGAAAGAGATGATTCGGGAACGTTAACAGCGCCAGGGGCTTTTCAATG 250	
0y	241	atcgaatcgaatgaagcttcgctgtgagtcacatgcccacgaacagtcagttgttcagag 300	

Db 251 ATGATGCGATGAAGGCTGCCGTTGAGTCAGATGCCCCAGCAAGTCAAGTGTGCGAGAC 310  
Oy ctgctactactagctgcgcaacagagcgctgactcttgcaagcgagaccgtctctgagagtg 360  
Db 311 CTGCTGACTATACCTGCGCAACAGACGCTACTCTTGCAAGCGGACCGCTCTTGAGAGTG 370  
Oy 361 ccgctcgctgcagcgtgactccctcctcaagcaatcctagatctggccaagcccaacttgct 420  
Db 371 CCGCTCGGTGACGTCAGTCCCTCAACAGGCAATGCTAGATCTGGCCCAACGCCCACTTGCC 430  
Oy 421 gctccattcttcaacctgccccagctggaagatagctttagaagcgtggcttgatcg 480  
Db 431 GCTCCATTTCTTCCCTGCCCGCCAGCTGAAGATAGCTTTAAGAAACGTGGGTCTGAATCGC 490  
Oy 481 tcgaagtgacctggtgctctgctccgagagacacacatttgaaagaaccagtgtagtgc 540  
Db 491 TCGAGTGACCTTGTGGCTCTGCTCCGAGAGACACACATTTGGAAAGAACCGTGTAGGTTTC 550  
Oy 541 atcatgatalagcgtctacaatttcaagcaaacactgggttlaactgaccaccaagctgaaact 600  
Db 551 ATCATGATATAGGCTCTACATTTTCAGCAACACATGGGTTACTTGACCCCAAGCTGAGACACT 610  
Oy 601 acgtatctccagacactgagagcgtgtgctcccaactgaatggcaacctcaagtactagtg 660  
Db 611 ACCTATCTCCAGACACTGAGAGGCTTGCCCACTGATGCAACCTCACTGCACTAGTG 670  
Oy 661 gacttgatctgaggcccccaaccatcttcgataaagaactatgtgaatcttagagagag 720  
Db 671 GACTTTGATCTGCGGACCCCAACCATCTTTCGATACACAGTACTATGTGATCTTAGAGAGAG 730  
Oy 721 cagaagagcctgaatacagaagtgaataagaactgtttagcagtlcccaagcaccactgaacc 780  
Db 731 CAGAAAGGCGCTGATACAGAGTGAATCAAGAACTGTTAAGCATGCCAAAGCCCACTGACACC 790  
Oy 781 atcccaactggtggaagcttctgtaactctaccctcaaacctctttaaagcctctgagaa 840  
Db 791 ATCCCACTGGTGAAGAGTTTGTCTACTACTACCAAACTTCTTTTAAACGCTTCGTGGAA 850  
Oy 841 gccatgaccgtatggtgaatacattaccctctgaaggttaaccgaagccagatctgctg 900  
Db 851 GCCATGGACCGTATGGGTATCACTTACCCTCTCTACCGGTACCCAGGCCAGATTCTGCTG 910  
Oy 901 aactgcagagtgtgtaacagcaactct 927  
Db 911 AACTGCAGAGTGTCAACAGCACTCT 937

RESULT 4  
A30294 1072 bp DNA PAT 09-OCT-1995  
LOCUS A30294 Synthetic IGF-II/HRP fusion protein gene.  
DEFINITION A30294  
ACCESSION A30294  
VERSION A30294.1 GI:1247537  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 1072)  
AUTHORS IGF-II ANALOGUES  
TITLE Patent: WO 93/03152-A 9 18-FEB-1993;  
JOURNAL Location/Qualifiers  
FEATURES  
source 1..1072  
gene /organism="Homo sapiens"  
/db\_xref="taxon:9606"  
misc\_feature 11..856  
11..856  
/gene="HRP"  
/partial  
/gene="HRP"  
CDS 11..1060  
/codon\_start=1

/product="IGF-II/HRP fusion protein"  
/protein\_id="CAA02021.1"  
/db\_xref="GI:1247538"  
/translation="MOTPTPEYDINSPNYSNIVRDTIVNELRSDPRIASITLRHFHD  
CFVNGDASITLIDNTTSFTEKDAQSNANSAAGFPVIDMKAAVSACRYSADL  
TIAAQQSVTLAGPSWRVPLGRSDLSQAELDLANMLPAPEFTLLPOLKSPFNGLNR  
SSDLVALSGHWFGRKNOGRIMDRLYNESNTGLPPTLTYTLQTLRGICPLNGNLSA  
LVDFDLRPTTFIDNKRYVNFLEOKGLIOSDELSSPNATDTIPIVRSFANSSTGFEN  
AFVEMAYRPSFTLCGGELVDITLQFVCGDRGRFYFSPRPSRVSRNSGIVEECFNSCD  
LALLETGYCATPKASE"  
misc\_feature 857..1057  
/partial  
/gene="IGF-II"  
gene 857..1057  
/gene="IGF-II"  
BASE COUNT 259 a 304 c 255 g 254 t  
ORIGIN  
Query Match 91.0%; Score 843.8; DB 5; Length 1072;  
Best Local Similarity 99.8%; Pred. No. 2,1e-245;  
Matches 845; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Oy 1 atgcagltaacccctacatctcagacaatagctgtcccaagctgtccaaactcgttcgc 60  
Db 11 ATCAGTTAACCCCTACATTCATGACAAATAGCTGTCCCAACGTGCCAATCATGTTCCG 70  
Oy 61 gacacaatgctaaagagctcagatccgatccagatgcgtctccaataatagctgc 120  
Db 71 GACCAATCGTCAACGAGCTCAATCCGATCCGATCCGATCCCTCTTCAATATACGCTG 130  
Oy 121 cacttcacatgactgctcgtgaaatgtgttgagcgtcagcatatctactggaacacacc 180  
Db 131 CACTTCCATGACTGCTTCGTAATGCTGTGGAGCGCTAAGCATATTACTGGCAACACCACC 190  
Oy 181 aglttcgcactgaaaaagatgcatctcgggaacgctaaacagcgccagggcttccagtg 240  
Db 191 AGTTTCGCGACTGAAGAGATGATTCGGGAAACGCTAACAGCCGAGGGCTTCCAGTG 250  
Oy 241 atcgatcgcatagaagctgcgctgttgagtgcaagcatgcccagaacagtctgtgcaag 300  
Db 251 ATCGATCGCATGAAGGCTGCGGTTGAGTCAGATGCCACAGCACTCACTTGTGAGAC 310  
Oy 301 ctgctactactagctgcgcaacagagcgctgactcttgcaagcgagaccgtctgagagtg 360  
Db 311 CTGCTGACTATACCTGCGCAACAGACGCTACTCTTGCAAGCGGACCGCTCTTGAGAGTG 370  
Oy 361 ccgctcgctgcagcgtgactccctcctcaagcaatcctagatctggccaagcccaacttgct 420  
Db 371 CCGCTCGGTGACGTCAGTCCCTACAGGCAATTCCTAGATCTGGCCAAAGCCCAACTTGCGCT 430  
Oy 421 gctccattcttcaacctgccccagctggaagatagctttagaagcgtggcttgatcg 480  
Db 431 GCTCCATTTCTTCCCTGCCCGCCAGCTGAAGATAGCTTTAAGAAACGTGGGTCTGAATCGC 490  
Oy 481 tcgaagtgacctggtgctctgctccgagagacacacatttgaaagaaccagtgtagtgc 540  
Db 491 TCGAGTGACCTTGTGGCTCTGCTCCGAGAGACACACATTTGGAAAGAACCGTGTAGGTTTC 550  
Oy 541 atcatgatalagcgtctacaatttcaagcaaacactgggttlaactgaccaccaagctgaaact 600  
Db 551 ATCATGATATAGGCTCTACATTTTCAGCAACACATGGGTTACTTGACCCCAAGCTGAGACACT 610  
Oy 601 acgtatctccagacactgagagcgtgtgctcccaactgaatggcaacctcaagtactagtg 660  
Db 611 ACCTATCTCCAGACACTGAGAGGCTTGCCCACTGATGCAACCTCACTGCACTAGTG 670  
Oy 661 gacttgatctgaggcccccaaccatcttcgataaagaactatgtgaatcttagagagag 720  
Db 671 GACTTTGATCTGCGGACCCCAACCATCTTTCGATACACAGTACTATGTGATCTTAGAGAGAG 730  
Oy 721 cagaagagcctgaatacagaagtgaataagaactgtttagcagtlcccaagcaccactgacac 780

Db 731 CAGAAAGCCCTGATACAGAGTATCAAGAACTGTTAGCAGTCCAAAGCCCACTGACACC 790  
QY 781 atcccaatggtgagaagttttgttaacttaactcaaccttcttaacgcttctgga 840  
|||||  
Db 791 ATCCCACTGGTGAAGAGTTTGTCTACTACTCAAACTCTTTAAACGCTTGTGGAA 850  
QY 841 gccatgg 847  
|||||  
Db 851 GCTATGG 857

RESULT 5  
AR000440  
LOCUS AR000440 1072 bp DNA PAT 04-DEC-1998  
DEFINITION Sequence 9 from patent US 5736363.  
ACCESSION AR000440  
VERSION AR000440.1 GI:3962971  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1072)  
AUTHORS Edwards,R.Mark and Bawden,L.  
TITLE IGF-II analogues  
JOURNAL Patent: US 5736363-A 9 07-APR-1998;  
FEATURES Location/Qualifiers  
source 1..1072  
BASE COUNT 259 a 304 c 255 g 254 t  
ORIGIN

Query Match 91.0%: Score 843.8; DB 5; Length 1072;

Best Local Similarity 99.8%: Pred. No. 2..1e-245;  
Matches 845; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgcagtaaaccttaactcttaagcaatagctgtcccaagctgtccaacatcgttcgc 60  
|||||  
Db 11 ATGCAGTTAACCCCTCATCTACATCTAGCAATAGCTGTCCCAAGCTGTCAACATCGTTCCG 70  
QY 61 gacacaatcgtcaagagctcagaatccgataccagagatcgctgtccaatataactctg 120  
|||||  
Db 71 GACACAATCGTCAACAGAGTCAGATCCGAGATCGCTTCATTAATATGCTCTG 130  
QY 121 caattccagtaactgctctgtaagtgttcgacgcttagcatattacttgaacaacacc 180  
|||||  
Db 131 CACTTCATGACACTGCTTCTGTGAATGTTGCGACGCTAGCATATTACTGGACACACCACC 190  
QY 181 agtttcgcgaactgaaaggatgcatcgcggaacgctaacagcgccaggggctttccagt 240  
|||||  
Db 191 AGTTTCGCGACTGAAGAGTGCATTGCGGAACGCTAACAGCGCGCTTTCAGTG 250  
QY 241 atcgatcgatgaaggctgcgttgaatgaatgcatgccccacgaacatcagttgtcaag 300  
|||||  
Db 251 ATCGATCGATGAAGGCTCCCTGTGAGTCAGCATGCCACAGCACTCACTTGTGCAGAC 310  
QY 301 ctgcataactataagctgcgcaacagagctgtaactcttcgagcgagacgctctgagaagt 360  
|||||  
Db 311 CTGCTACTATAGCTCGCAACAGACGCTGACTCTTGCAGCGGACGCTCTGGAAGTG 370  
QY 361 ccgctcgtcgtaagtgactccctactacagcatctcctagatctgccaacgccaacttgct 420  
|||||  
Db 371 CGCCTCGTGCGAGTGAACCTCTACAGGCAATTCCTAGATCTGGCCAACGCCCACTTGCC 430  
QY 421 gctccattcttcaacctgccccagctggaagatagctttagaagatggtgtcgaatcgc 480  
|||||  
Db 431 GCTTCATTTCTTCACCTGCGCCAGCTGAAGGATAGCTTTAGAAACGTGGGTTCTGAATCGC 490  
QY 481 tcgaagtaacttggctgtctgtccgagagacacacatttggaaagaacagtgtaagtttc 540  
|||||  
Db 491 TCAGATGACCTTTGGCTGTGTCGGAAGACACACATTTTGGAAAGAACAGTATAGTTTC 550  
QY 541 atcatgatagtgctctacaatttcagcaacactgggttactctgaccccaagctgaacact 600

Db 551 ATCATGGATAGGCTCTACAAATTTTCAGCAACACTGGGTTACTGACCCCAAGCTGAACACT 610  
QY 601 acgatctcagaacactgtgagaagctgtgtgcccaactgaatggaacctcagtgtaactagtg 660  
|||||  
Db 611 ACCTATCTCCAGACACTGAAGAGGCTTGTGCCACTGAATGGCAACCTCACTGACATAGTG 670  
QY 661 gacttgatctgcgagccccaacatcttcgaacaagaactatgtgaatctagaaggag 720  
|||||  
Db 671 GACTTTGATGTGGGAGCCCAACCATCTTGATACAGACTATGTGAATCTTAGAGAG 730  
QY 721 cagaagggcctgatacagagtgatacgaactgttttagcagtlccagaagcgaactgacacc 780  
|||||  
Db 731 CAGAAAGCCCTGATACAGAGTGAATCAAGAACTGTTAGCAGTCCAAAGCCCACTGACACC 790  
QY 781 atcccaatggtgagaagttttgttaacttaactcaaccttctttagagcttctgga 840  
|||||  
Db 791 ATCCCACTGGTGAAGAGTTTGTCTACTACTCAAACTCTTAAACGCTTGTGGAA 850  
QY 841 gccatgg 847  
|||||  
Db 851 GCTATGG 857

RESULT 6  
AR068580  
LOCUS AR068580 1072 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 9 from patent US 5854025.  
ACCESSION AR068580  
VERSION AR068580.1 GI:6000787  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1072)  
AUTHORS Edwards,R.Mark and Bawden,L.  
TITLE IGF-II analogues  
JOURNAL Patent: US 5854025-A 9 29-DEC-1998;  
FEATURES Location/Qualifiers  
source 1..1072  
BASE COUNT 259 a 304 c 255 g 254 t  
ORIGIN

Query Match 91.0%: Score 843.8; DB 5; Length 1072;  
Best Local Similarity 99.8%: Pred. No. 2..1e-245;  
Matches 845; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgcagtaaaccttaactcttaagcaatagctgtcccaagctgtccaacatcgttcgc 60  
|||||  
Db 11 ATGCAGTTAACCCCTCATCTACATCTAGCAATAGCTGTCCCAAGCTGTCAACATCGTTCCG 70  
QY 61 gacacaatcgtcaagagctcagaatccgataccagagatcgctgtccaatataactctg 120  
|||||  
Db 71 GACACAATCGTCAACAGAGTGCAGATCCGAGATCCGCTGCTTCATTAATATGCTCTG 130  
QY 121 caattccagtaactgctctgtaagtgttcgacgcttagcatattacttgaacaacacc 180  
|||||  
Db 131 CACTTCATGACACTGCTTCTGTGAATGTTGCGAGCTAGCATATTACTGGAACACACCACC 190  
QY 181 agtttcgcgaactgaaaggatgcatcgcggaacgctaacagcgccaggggctttccagt 240  
|||||  
Db 191 AGTTTCGCGACTGAAGAGTGCATTGCGGAACGCTTAACAGCGCCCAAGGCTTTCAGTG 250  
QY 241 atcgatcgatgaaggctgcgttgaatgcatgccccacgaacagtlgttgcagac 300  
|||||  
Db 251 ATCGATCGATGAAGGCTGCGTTGAGTCAAGTCCACAGAACAGTCACTGTGACGAC 310  
QY 301 ctgcataactataagctgcgcaacagagctgtaactcttcgagcggaacgctctgagaagt 360  
|||||  
Db 311 CTGCTACTATAGCTCGCAACAGAGGAGTGAATCTTGCAGCGGACGCTCTGGAAGTG 370

QY 361 ccgcctcgctcgacgctgactcccttaagcacttcctagatcctgagccacgcgcaacttgct 420  
|||||  
Db 371 CCGCTGGGTGAGAGTACTCCCTACAGGCAATTCCTAGATCTGGCCACCACTTGCCCT 430  
QY 422 gctcaattcttaacccctgcccaggaagatagctttagaagcgtggctgaaatgc 480  
|||||  
Db 431 GCTCCATCTCTACCCCTGCCCCAGCTGAAGGATAGCTTTAAGAAACGTGGGTGAAATCCG 490  
QY 481 tcgaatgactctgctgctcctcgagagacacacatttgaaagacagtgtaagtc 540  
|||||  
Db 491 TCGAGTACCTTGTGGCTCTGTCGGAGGACACACATTTGGAAAGAACAGTGTAGGTTTC 550  
QY 541 atcaatgataagctctacaatttcagcaaacactgggttaacctgaagcccaagctgaacact 600  
|||||  
Db 551 ATCATGGATAGGCTCTACATTTTACCAACACACTGGGTTACTCTGACCCCAAGCTGAACACT 610  
QY 601 acgatcttcagacacttgagagctgtgtccactgaataggcaacttcagctgactagtg 660  
|||||  
Db 611 AGCTATCTCCAGACACTGAGAGGCTTGTGCCCACTGAATGGCAACCTCAGTGCATAGTG 670  
QY 661 gactttgactcgagagcccaacacatcttcgataacaagtaactatgtgatatcagaagag 720  
|||||  
Db 671 GACTTTGATCTGGGGAGCCCAACACATCTTCGATTAACAAGTACTATGTGAATCAGAGAG 730  
QY 721 cagaaagcctgatacagagtgatacaagaactgttttagcagtcacagacgcacacacac 780  
|||||  
Db 731 CAGAAAGGCCCTATACAGAGTGTCAAGAACTGTTTACAGTCCAAACGCCCACTGACACC 790  
QY 781 atcccactgtgagaaagtttcttaacttactcaaaccttcttaacgcttcgtgaa 840  
|||||  
Db 791 ATCCCACTGGTGAGAAAGTTTGTCTAATCTACTCAAACTCTTTTAAAGCCCTTGTGTGA 850  
QY 841 gccatag 847  
|||||  
Db 851 GCTATGG 857

RESULT 7  
A06250  
LOCUS A06250 927 bp DNA  
DEFINITION Synthetic HRP isoenzyme C gene.  
ACCESSION A06250  
VERSION A06250.1 GI:1566731  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
1. 927  
/db\_xref="taxon:32630"  
BASE COUNT 247 a 263 c 198 g 219 t  
ORIGIN

Query Match 60.9%; Score 564.6; DB 5; Length 927;  
Best Local Similarity 75.7%; Pred. No. 1.5e-160;  
Matches 699; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 4 cagtttaaccccttaacttcaagacatagctgtcccaagtggttccaactcgttgcgac 63  
|||||  
Db 1 CAGTTTAAAGCCCACTTTCTACAGTAAACACTCTCCAAAGCTCTCGATATTTGTCCCGAT 60  
QY 64 acaatgcataaagactcagatccgataccagatgcgtccttcaatatatcgtcagc 123  
|||||  
Db 61 ACCATCGTAATAGAGCTGGGTCCGATCCACGTATTGACAGTACTCTTGTCTACTAT 120  
QY 124 ttccatgactgctcgtgaatggttcgcagcgtctagcatactctgagacaacaccagct 183

Db 121 TTTTCATGACTCGCTTTTAAACGGATGTGACGGGTCCATCTCTCTGGAACAATCAACAAC 180  
QY 184 ttccgcagtaaaaggaatgatttcgggaacgcttaaaagcgaagggcttttcagtgatc 243  
|||||  
Db 181 TTCCGAACTGAGAAAGACCATTCCTGGGAACGCAATTCGGCCCGGGCTTTTCCAGTCATC 240  
QY 244 gatcgatgaagctgcgcttgatgataagtcaccaagacagtcagtgctgcaacctg 303  
|||||  
Db 241 GATCGCATGAAGGCTTCATTCGATCGGCTCTGCCGAGAACTGTTTTCGTGTGACAGCTG 300  
QY 304 ctgactatagtctgcgcaacagagcgtgactcttgcaagcggagacgctcctggaagtgccg 363  
|||||  
Db 301 CTGACCATGCGAGCCCAAGCAATCGGTCACCTTACCGCGGGGCCCTCATGCGGAGTACCA 360  
QY 364 ctcggtcgagcgtgactcccttaagagatctcctagatctggtccaagccaacttgctgt 423  
|||||  
Db 361 TTAGGAGCTGTGACTCCCTTACAGGCTTTCTTGAATCTAGATCTGAATGCAATCTTACCGGCG 420  
QY 424 ccattctcaccctgcccagctgaaagataagctttagaacgctgggtcgtgaactcgtg 483  
|||||  
Db 421 CCATTTTACTCTACCTACCAACTTAAGATTCATTTTCGTAGCTAGGCTTGAAACGCTCG 480  
QY 484 agtgacctgtgctctgtcgcggaaggaacaacatttgaaagaaacagtgtagttcatc 543  
|||||  
Db 481 AGTATCTAGTGTGCTCTTCTGTGTGGCCACACTTTTGGAAAGAACCAAGTGCATTCATC 540  
QY 544 atggaatgactcttaacatttcaagcaacatgggttaacctgcaagccagctgaaacctag 603  
|||||  
Db 541 ATGGACCGATTTGTACACTTTTTCAAACACTGACTGACCGGATCCAGCTCAATAGAGACC 600  
QY 604 tatctcagaacactgaaagactgtgtcccaactgaatgtaaacctcagctgactagtgac 663  
|||||  
Db 601 TACCTACACACACTCCGGGACTATGCCCTTTAAAGGAACCTTAAGCCGCTAGTCGAC 660  
QY 664 ttgatctgcggaaccccaacalcttgataaacaagtactatgtgaatagaagagag 723  
|||||  
Db 661 TTTGACCTACGTACGCGCAACAATCTTCGACAACAAGTACTATGTAAATCTAGAGAACAA 720  
QY 724 aaaagcctgatacagaatgatacaagaactgtttagcaagtcagcaagccactgaacccatc 783  
|||||  
Db 721 AAAGCATTAATTCATTCATGATCAGAGCTTCTCTCATCCCAATGCAACAACAGATC 780  
QY 784 ccactgtgtagaagtttcttaacttactcaaaccttctttaaagccttcgtggaagcc 843  
|||||  
Db 781 CCACGTAGTGTGATGCTTTCGAACTCAACAAACCTTTCTTAATGCAATTCGTAGAGGCC 840  
QY 844 atggacgtaatggtgatacaattacccctctgacggtgataccaagccagatcgtctgaac 903  
|||||  
Db 841 ATGGATCGAATGGAATATATCACGCCACTYCACGGGTACCAAGTACAGATCAGTGAAC 900  
QY 904 tgcagagtggtcaacaagaacac 926  
|||||  
Db 901 TCCCGAGTGTAAACTCGAATCTC 923

RESULT 8  
E01651  
LOCUS E01651 924 bp RNA  
DEFINITION cDNA encoding horseradish peroxidase.  
ACCESSION E01651  
VERSION E01651.1 GI:2169904  
KEYWORDS JP 1988207386-A/1.  
SOURCE  
ORGANISM  
Armoriacia rusticana.  
Armoriacia rusticana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Caprales; Brassicaceae; Armoriacia.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Okada, H., Niina, A. and Yamada, Y.  
STRUCTURAL GENE OF HORSERADISH PEROXIDASE  
Patent: JP 1988207386-A 1 26-AUG-1988;  
TOYOBO CO LTD

COMMENT	OS (horseradish)
PN	JP 1988207386-h/1
PD	26-AUG-1988
PF	19-FEB-1987 JP 1987037623
PI	OKADA HIROSHI, NIINA ATSUSHI, YAMADA YASUYUKI, PC
	CI2N1/00, CI2N1/20, CI2N9/08, (CI2N1/20, CI2R1:19), (CI2N9/08, PC
	CI2R1:19);
CC	strandedness: Double;
CC	topology: Linear;
CC	hypothetical: No;
CC	anti-sense: No;
CC	*source: strain=Armoracia lapathifolia, Armoracia rusticana;
CC	*source: tissue-type=root;
CC	*source: clone=PSK1;
CC	feature is identified by similarity;
CC	key Location/Qualifiers
FT	FT
FT	FT
FEATURES	Location/Qualifiers
Source	1..924
	/organism="Armoracia rusticana"
	/db_xref="taxon:3704"
BASE COUNT	256 a 223 c 188 g 257 t
ORIGIN	
Query Match	56.5%; Score 524; DB 5; Length 924;
Best Local Similarity	72.9%; Pred. No. 3,2e-148;
Matches 674; Conservative	0; Mismatches 250; Indels 0; Gaps 0;
OY	4 cagtaaccctacatctcagacaaatagctgtcccaacgtgtcccaacatcgttcgcgac 63
DB	1 CAACCTACCCCTACCTCTTACGACAACTTCATGTCTTAATGCTTAACATCGTACG 60
OY	64 acaatcgtcaacagagcttagatccgatcccaagatcgcgtgtcctaataatagcttgac 123
DB	61 ACTATTTGCAATGACCTTAAGATCAGACCCCTGATTTGCCGCGACGATCCCTTCCTAC 120
OY	124 ttccaatgcgtctcgttgtaatgtgttcgcagcgttagcatalatctggaacaacacacagt 183
DB	121 TTCCAGCAGCTGCTTTGTTAATGTTGTGACGAGATGATCTTGTACACACAACTATCA 180
OY	184 ttccgactgaaagagatgcatctcggaacgtcaacagcgccagggctttccagtgatc 243
DB	181 TTTTCGACAGAGAAAGATGCGTTTGGAAACGCAACTCGGCAAGAGGATTTCCAGTAT 240
OY	244 gatcgcatgaagagctgtccgttgatgcagcaatgcccaagatcagttgtgcgaactgt 303
DB	241 GATTAATAATGAAGCGCGCGGTGAGAGTGCATGCCCAAGAACCGTTTCATGCCACGATTTTG 300
OY	304 ctgactatagcttgccgaagaagcgttgactcttgcaagcgagccgttcctggagagtgccg 363
DB	301 CTCACCATTTGCAAGCTCAACAATCTGTCACTTTGGCGGGAGGTCTTCTTGAGAGTTCT 360
OY	364 ctccggtcgaagtgatccctcctcaacagcattcctctaga tctgtgccaaagcccaactgtcgtc 423
DB	361 TTGGCGCAGAGAGATGAGCTTACAAACATTTTCGGATCTTGCTAATGCAAAATTTCCAGT 420
OY	424 ccaatcttcaaccctgcccagctgaagatagctttagaacgtgtgtgttcgaatcgtctg 483
DB	421 CCATTCCTTCACACTTCCCAACAATCTTAAAGACACTTTAGAAATGTGGCCTCAACCGTTCT 480
OY	484 agtgacgtctgtggtcctcttcggagaagaacacattggaagaagaacagtggtgttcac 543
DB	481 TCTGATCTGCTGTGCACTGTCCGGGGCCACACATTTTGGTAAAAAATCAGTGTGGTTTAT 540
OY	544 atggaatgagctctaaacttcagcaaacactgtgttactcgtgaaccccaagctgaacactacg 603
DB	541 ATGGACAGATTATACACTGTCAGCAACCGGTTTACCGGATCTTACTGTCAACACTACT 600
OY	604 tatctccagaacatgagaggtctgtgcccaatgaaatggaacatcagttcagttactagtgac 663

Db	601	TATCTCCAAACTCTTTGGGACACTATGTCCTCCATATGTAATCTAAGGCGCTTTGGTGCAT	660
Qy	664	ttgatctgcggaccocccaccatcttcgataaacaagtactgtgaatctagaagagacag	723
Dg	661	TTTGATCTACCTAGCGCCACMGATTTTGTGACAAACAATAACTATGTGTAATCTGAAAGAGAA	720
Qy	724	aaagacctgatacagagatgatacaagaactgtttcagctccagagccatctgaaccatc	783
Db	721	AAAGACTTATCCAAAGGACCAAGAGAGTGTCTTCACCCCAATGCGCATGACACATTC	780
Qy	784	ccactgtgagagaagttttgctactactactaaaccccttcatacgcctctgtggaacg	843
Db	781	CTTTGGGAGATATCATTTTGGTAAATAGCACACAAACATTTCTTCAATGCAATTTGTGGAGCG	840
Qy	844	atggaccgatacgttgataacattaccctctcgaagggtacccaaggcccaatctgctcgaac	903
Db	841	ATGGATAGATAGGGAACACATTATACACTCTTACAGGAATCAAGGACAGATCAGGTGAAT	900
Qy	904	tgcagagtgtgcataacagcaactct	927
Db	901	TGTAGGGGTGAGTAACCTCACTCT	924
RESULT	9		
LOCUS	ATPRXCB	1236 bp	PLN
DEFINITION	A. thaliana prxcb mRNA for peroxidase.		04-NOV-1994
ACCESSION	X71794		
VERSION	X71794.1 GI:405610		
KEYWORDS	basic peroxidase; heme; prxcb gene.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryotes: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.		
AUTHORS	1 (bases 1 to 1236)		
TITLE	Intapruk,C., Takano,M. and Shimmyo,A.		
JOURNAL	Nucleotide sequence of a new cDNA for peroxidase from Arabidopsis thaliana		
REFERENCE	Plant Physiol. 104 (1), 285-286 (1994)		
AUTHORS	94159792		
TITLE	2 (bases 1 to 1236)		
JOURNAL	Shimmyo,A.		
COMMENT	Direct Submission		
FEATURES	Submitted (30-APR-1993) A. Shimmyo, Osaka University, Dept of Biotechnology, Faculty of Science, Suita, Osaka 565, JAPAN		
source	Related sequences: Intapruk,C. et al., Gene. 93:237-241 (91) & Intapruk,C. et al., J. Ferment. Bioeng. 75:166-171 (93).		
location/Qualifiers			
1..1236			
/organism="Arabidopsis thaliana"			
/strain="L. Columbia"			
/sub_strain="wild type"			
/db_xref="taxon:3702"			
/tissue_type="whole plant"			
35..1096			
/gene="prxcb"			
35..1096			
/gene="prxcb"			
/EC_number="1.11.1.7"			
/codon_start=1			
/product="peroxidase"			
/protein_id="CA50677.1"			
/db_xref="GI:405611"			
/db_xref="SPTREMBL:Q42564"			
/translation="MHPSSSTSTWTLITLGLMLHSLSAQLTPFYDRSCPWV			
TNIYRETIYVDELSDPRILIASILRLHHDCEVNGCDASILLDTTSPTREKDRGNAN			
SAREPITIDMKAAVERACPRVSCADMLTIAOOSVTLAGGSPMRVPLGRDLSIAF			
MLGANALPAPEFTLPOLKASPRVNGCDRPSDLVALSGGHTFGNOCQFLIDRPYNS			
TEGLPDLTMTYTIQTLRGCLPLNGNSALVDDPURLPYGFNDKTYVNLKEKGLIOS			
DOELFSSPNATDITPLVRAIADGQTGFENAFVEAMNRMGNTPTTGOGIILNCRAV			
NSNSLHLDVYDIDVEVSSM"			
1144..1149			

```

BASE COUNT      376 a      282 c      229 g      349 t
ORIGIN
/notes="basic subunit"
Query Match      53.2%; Score 493.6; DB 7; Length 1236;
Best Local Similarity 70.9%; Pred. No. 5.7e-139;
Matches 655; Conservative 0; Mismatches 269; Indels 0; Gaps 0;

QY 4 caattaccctcatctctcagcaatagctgtcccaagctgtcccaactcgttcgagac 63
DB 125 CAACCTACCCCTTACCTTTACGATAGTCTATGCTTATCTACATACATGACGAGAA 164
QY 64 acaatcgtcaacagctcagatccgaatcccaagatcgtctcattatctacgtcagc 123
DB 185 ACCATTGTAATAGATTAAAGCTGGAGCCCTGATGCCCTGGAGCATCCCTGCTTAC 244
QY 124 ttccatgactgctcgttgatggtgtgcagcgttagcatatctactgcaaacaccagc 183
DB 245 TTCACAGACTGCTTTGTTAATGGTTGGACGATCCATCTTGTAGCAACACGACATCA 304
QY 184 ttccgacatgaagaagatgcatcgcggaacgctaacagcgccggggttcgaatgac 243
DB 305 TTTCGACAGAGAGAAAGATCGGTTTGGAAACGCAAAATTCGGCTCGGGGATTTCCAGTGAT 364
QY 244 gatcgcatgaagagctgcgttgatgacatgcccacgaaacagtcagttgtgcagactg 303
DB 365 GATAGATGTAAGACTCGGGTGGAGAGGCGATGCCCAAGACCGTTTCATGCCAGATATG 424
QY 304 ctgacatagctgtgcgaacagagcgtgactctgcagcgagcgacgtccgtggaagtgcg 363
DB 425 CTCACATTCGACGCTCAACATCTGTCACTTTGGACGAGGCTCTTGTGGAGGTTCT 484
QY 364 ctggtcgacgcttactccctacagcatctcctagatctgccaagcccaactgctgct 423
DB 485 TTGGGAAGAGAGACAGATTTCACAGATTCCTTGAACCTCCCTAATGCAAAATCTTCCAGCT 544
QY 424 ccaatctcaaccctgcgcaacgtgaagatagctcttagaaacgttggtgtcgaatcgtcg 483
DB 545 CCATCTTTACACTTCACACTCCACAACTTAAAGCCAGCTCAGAAATGTTGTCGATCGCT 604
QY 484 atgacactggtgctctgtgcggaagacaacatcttgaaagaacagtgtagttcatc 543
DB 605 TCGATCTGCTGCTCTCTCGGTGGTGCACACATTTGGTAAATGATAGTGTGAGTTATT 664
QY 544 atgcatagctctacaaatttcagcaacactggtgttaacctgaccccgctgaacactacg 603
DB 665 CTTCGACAGATTTCACAAATTTACAGAACACAGGTTTACCCGACCTTACACTCAACACTACT 724
QY 604 tatccagaacatgagagctgtgtgccatgaaatgcaactcagtgactagtgagac 663
DB 725 TACCTCAAACTCTTGATGAGATATGCCCTTAATGGCAATGAGATGCTTGGTAGAT 784
QY 664 ttgactctgcggaagcccaacatctcgtataaagtaactatgtatgtatagagaagcag 723
DB 785 TTTGATCTACGTAAGGCTTACGGTTTTCGACAAACAAATACTACGTGATCAACAAAGCGCA 844
QY 724 aaagcctcatatagagatgatacaagaactgtttagcagttccagagcccaactgacacatc 783
DB 845 AAAGGCTTATTCAGAGCGCAACAAAGTTGTCTTACAGCCCAATGCGCATGACAAATC 904
QY 784 ccaatggtgagaagatttgccttaactcctcaaaccttctttaaagcgttcgtlgaagcc 843
DB 905 CCCCTGTGAGAGCATATGCTGATGGCACACAACAAATTCCTCAATGCATTTTGGAGGCA 964
QY 844 atggaacgtatggtgataacttaaccctctgacggtgtacccaagccagatgtctggaac 903
DB 965 ATGAATATGATGGGAAACATTTACACCACTACAGGAATCMAAGACAAATACAGATTGAAC 1024
QY 904 tgcagagtggtcaacagcaactct 927
DB 1025 TGTAGAGTTGTGATCACTCAACTCT 1048

```

```

RESULT 10
HRAHRPCC      1229 bp      DNA      PLN      27-APR-1993
LOCUS      A.rusticana peroxidase isoenzyme C (HPR C) gene, complete cds.
DEFINITION      M60729
ACCESSION      M60729.1 GI:168244
VERSION      horseradish peroxidase.
KEYWORDS      horseradish peroxidase.
SOURCE      A.rusticana DNA, clone prxc3.
ORGANISM      A.rusticana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Armoracia.
REFERENCE      1 (bases 1 to 1229)
AUTHORS      Fujiyama,K., Takemura,H., Shibayama,S., Kobayashi,K., Choi,J.-K.,
Shimizu,A., Takano,M., Yamada,X. and Okada,H.
JOURNAL      Structure of the horseradish peroxidase isozyme c genes
MEDLINE      Eur. J. Biochem. 173, 681-687 (1988)
FEATURES
source      1..1229
              location/Qualifiers
              1..169
              /organism="Armoracia rusticana"
              /db_xref="taxon:3704"
              1..169
              /gene="HRPC3"
              /number=1
              join(2..169,170..361,362..530,531..1000)
              /gene="HRPC3"
              join(2..169,170..361,362..530,531..1000)
              /gene="HRPC3"
              /codon_start=1
              /protein_id="AAA33379.1"
              /db_xref="GI:168245"
              /translation="MHASFSAQLPTFFDSCSRVSNIVDIIINELRSPSTIAS
              ILRLHFHDCFVAGCDASILLDNFTSFRLEKDAFGNANSARGPVVDRIKAAVERACP
              TVSCADVLTLIAQOSVNLALGSPSRVPLERDSRQAFDLANANLPAESFTLPELKA
              FANVGINRPSDILVALSGHTFEGNOCREFIMDRILYFNSGTPLDPTLNTYTLQTLKQC
              PENCNOSYLVDLRLTPVFDNRKYVNLKEQGLIOSDELFSFNATDTIPLVSYA
              DETQFFPNAFVAMRMGMNIPPLUTGTGSRILNCRVYVNSILHIVEVDVSSM"
              170..361
              /gene="HRPC3"
              /number=2
              362..530
              /gene="HRPC3"
              /number=3
              531..1000
              /gene="HRPC3"
              /number=4
BASE COUNT      337 a      275 c      256 g      361 t
ORIGIN
Query Match      52.9%; Score 490.4; DB 7; Length 1229;
Best Local Similarity 70.7%; Pred. No. 5.4e-138;
Matches 653; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

QY 4 caattaccctcatctctcagcaatagctgtcccaagctgtcccaactcgttcgagac 63
DB 29 CAACCTACCCCTTACCTTTACGATAGTCTATGCTTATCTACATACATGACGAGAA 88
QY 64 acaatcgtcaacagctcagatccgaatcccaagatcgtctcattatctacgtcagc 123
DB 89 ATCATATTACAAATAGTTACAGATGAGACCCGATGATGCGCGGAGCATTCGTGCTTCCAC 148
QY 124 ttccatgactgctcgttgatggtgtgcagcgttagcatatctactgcaaacaccagc 183
DB 149 TTCACAGACTGCTTTGTTAATGGTTGTGACGATCGATCTTGTAGCAACAAACATCA 208
QY 184 ttccgacatgaagaagatgcatcgcggaacgctaacagcgccggggttcgaatgac 243
DB 209 TTTCGACAGAGAGAAAGATCGGTTGGAACGCAACCTCGGCTCGAGATTTCTGTGTT 268

```



OY	244	gacgcagtagaggttcgctctgtgtagtcaagatgcccaagcaagctgaattgtgtcagactg	303
Db	269	GACGAATCAAGGCGCGGTGGAGGGGATGGCCCAAGAACTGTTTCAATGCGCAATGTG	328
OY	304	ctgactatagctgcgcaacagagcgtgactctctgcagcgagacgtctccgagagtagtcg	363
Db	329	CTTACCAATTGCACTCAACAATCTGTTAATTTTGGCAGGAGGTCTTCTTGGAGGGTTTCT	388
OY	364	ctcgtctgacgttactccctactacagcatctccactagatctgtgccaacgcgcaacttgcgt	423
Db	389	TTGGGAGAGAGACAGCGGACCAAGCATTTTTAGATCTCGTAATGCGAATTTCCACT	448
OY	424	ccactctcacccctgcgccagcgtgaagatagctttagaacygtgtgtctgtaactgcgtcg	483
Db	449	CCATCTCTTCAACATTCCAGAACTTAAGGCGTGGCTTTTGGAAATGTTGGCCTCAACGTCCT	508
OY	484	agtgacctgtgtctctgtccgaggaacacatcttggaaagaacagtgtagttatc	543
Db	509	TCTGATCTCTGCTCTCTCTGTGGTGTCAACATTTTGTAAATAATCAATGTCGATTTATTT	568
OY	544	atgatatgagctctacaatttcagcaaacacttgggttaccgaccccaagctgaacatcag	603
Db	569	ATGAGACGATTAATACAACTTCAGCAACACCGGACTACCGGACTTACCTCAACACTACT	628
OY	604	tatctccagacacttgagaggtctgtgcccactgaatgtgcgaacctcagtgcatagttgac	663
Db	629	TACCTTCAAACTCTTGCTCAACAATGTCCCGTAATGTAATCAACAAAGGCTTGGTGAT	688
OY	664	tttgatctgcggaacccaacalcctcagataacaagtaactatgtgaactctagagagcag	723
Db	689	TTTCATCTGGCTACGCCAACAGTTTTCGATATACAAATTAATGTAATCAATTTAAAGCAA	748
OY	724	aaaagccttgataaagtagtcaagaacggttagaggtccaggtccagcgcaactgaacacatc	783
Db	749	AAAGGTTCTTCCAGAGGTGACCAAGAGTGTCTTAGGCCCAATGACCACTGACCAATCTG	808
OY	784	ccaactgttgagaagtttctctaactactacaacctctttaaagcctctgttgaaacc	843
Db	809	CCCTTGTGAGATCATATAGTGTGATGGCACACAAACATTTTCAATGCTCTTGTGTGAGGCC	868
OY	844	atggaaccgatgtgtaacatcaccctctgacggtgtacccaagccagatctgcttgaaac	903
Db	869	ATGCAATGGAGTGGCAAAACATTTACACCTCTTTACAGCAATCAAGAGAAATCAGGTTGAAC	928
OY	904	tgcagagtggtcaacgcaactctc	927
Db	929	TGTAGGGTGTGAACTCCAACTCT	952

FEATURES	Location/Qualifiers
source	1..1195
	/organism="Arabidopsis thaliana"
	/cultivar="Columbia"
	/db_xref="taxon:3702"
	/tissue_type="stems, leaves, flowers, roots and siliques"
	/clone_1ib="Lambda PRL2 (Dr. T. Newman et al)"
	/clone="EST clone 91D22r7"
	32..118
sig-peptide	32..1090
CDS	32..1090
	/ec_number="1.11.1.7"
	/codon_start=1
	/product="peroxidase AtPp16a"
	/protein_id="CNA67313.1"
	/db_xref="GI:1429221"
	/db_xref="SPTRMBL:O43732"
	/translation="MNFSSYSLSHWYTLMTLGCILLHLSTSSAQLPTFYDNGPSVFTIYRVDIVNELRSDPRIAASILRLRHFDCCYNGCDASILDNNTTSFRTEDDAAPNANSARCPVADIRMKAAVETACPRVSCADILTLIAAQOVAAGSGSKRWPLGHRDSLQAFATLADNLPAPFTYLTLPQLKASFRONGIDRPSDLVLSGHTFGKNOCCFIMDRPLVNESTGTADDPDLNTTYTLPQTLTGQCPFRNNGNGLVDLDRTPYFDNKKYVNLKELKGLIQDQELSSPNADTQPIPLVREYADGQTKFPAALFEAANNRNGNITPLLTGGQILRONCRVYVNSNLLHIVETIVDVSSM"
mat-peptide	119..1042
BASE COUNT	354 a 278 c 223 g 340 t
ORIGIN	

RESULT 11	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
ATPEROX5	1195 bp	RNA	PLN	07-OCT-1996							
	A.thaliana mRNA for peroxidase AtPl6a, EST clone 91D2127.										
	X98777										
	X98777.1	GI:1429220									
	peroxidase										
	thale cress.										
	Arabidopsis thaliana										
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;										
	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;										
	Rosidae; Caprales; Brassicaceae; Arabidopsis.										
	1 (bases 1 to 1195)										
	Wolinder,K.G., Jespersen,H.M., Kjærsgaard,I.V.H., Justesen,A.F.,										
	Ostergaard,L., Abelskov,A.K., Jensen,R.B., Hansen,L.N. and										
	Rasmussen,S.K.										
	From expressed sequence tags to structure, function, evolution and										
	expression of 28 ER-targeted Arabidopsis peroxidases										
	unpublished										
	2 (bases 1 to 1195)										
	Justesen,A.F.										
	Direct Submission										
	Submitted (25-JUN-1996)										
	A.F. Justesen, Institute of Molecular										
	Biology, Department of Protein Chemistry, Øster Farimagsgade 2A,										

Query Match	51.9%	Score 481.4	DB 7	Length 1195
Best Local Similarity	70.1%	Pred. No. 2.9e-135		
Matches 647	Conservative 0	Mismatches 276	Indels 0	Gaps 0
QY 4	cagtaacccttacaattctacagacaatagctgltcccaacgltgccaaatcgtltcgagc	63		
Db 119				
	CAGTAACCCCAACTTCTATATGACAAATACATGTCCAAAGCGTCTTATCCATCGTAGCGGAT	178		
OY 64	acaatcgtccaagagcttagatccgatcccaaggatcgtgtcttcaatatcagcttcgac	123		
Db 179				
	ACCAATGTAACAGACTTAAGATCGGATCTCGTATCGCTGCAAGCATCTTTCGCTTAC	238		
OY 124	ttccatgactcgttcgttgaatgltgttcgagcctagacatatcttgcgaacaaccacagt	183		
Db 239				
	TTCACGACGCTGTTTCGCAATGGCTGTGATGCATGCATCTGTTATGACAACACCAATCC	298		
OY 184	ttccgcactgtaaaagatgcatcttcggaaagctaaacagcccaaggcttccagtgatc	243		
Db 299				
	TTTTCACACGAGAAAGATGCGGCTCCAAACCAACAACTCGCGTGGATTTTCCATGATTT	358		
OY 244	gatcgcatgaagagctgcgcgtttgatgcagcatgcgccaagcaagtcagtttgcagacctg	303		
Db 359				
	GATGAAATGMAAGCAGCAGTAGAGACACATGTCCAAAGACTGTTTCATGCGCAGATATA	418		
OY 304	ctgactatagctgycgaacagagcgltgactcttcggacgycgagccgttccatgagatgccc	363		
Db 419				
	CTTACCATTCGACGAGCAACAAGCTGTAATTTTGGCAGGAGGCTCTTCTTGGAGGGTTCCT	478		
OY 364	ctcgatgacgttgcactccctaagggcatcccttagatctgycgcaagccaacttgctctc	423		
Db 479				
	TTGGGAGAGAGAGAGCTTACAAACATCTTTTGGCTTCGCTAAATACAAATCTTCCGCT	538		
OY 424	ccatctctaccctgccccagctgtaagatagctttagaaacgttggctctgaatccgtcg	483		
Db 539				
	CCATCTTCACTTCCCACTTAAAGCCAGCTTTCAAAATGTGGACTTGACCGTCC	598		
OY 484	agtagaccttgggcctgttcggggagaaacacatttgaagaaccagttaggtttc	543		
Db 599				
	TCTGATCTCGTTGCACTCTGTGTGTGTAACAATTTGGTAAAAAACCACTGCCAATTTTAT	658		
OY 544	atggaatagcttcaaatctcagacaacatgggttattcttgaccacagctgaacatacag	603		
Db 659				
	ATGGATACATATACACTTTAGCAACAAGCTGTTTACCAGACCCATCTTAAATCTACA	718		



Db 835 AAGGACTTATTCAGACGATCAAGAGTTGTTTTCAGCCCTAATGACACGACATC 894  
QY 784 ccaatggtgagaagtttcttaactactcaaaccttctaagcgttcgtggaagcc 843  
Db 895 CCTTGTGTGAGACATATACCGTGTGACTGCAAAAATTTCTCAATGCTTTTGTGGAAGCA 954  
QY 844 atggaccgtatggtaacattaccctctgaagcgttaaccgaagccaagatctgctgaac 903  
Db 955 ATGAATAAGATGGGAACAATTACACTCTTCACGTGGAATCAAGACAAATACAGACGAAT 1014  
QY 904 tgcagagtggtcaacagcaactc 926  
Db 1015 TGTAGGTGTGTTAACTCAACTC 1037

RESULT 13  
PPXP2PER  
LOCUS PTPXP2PER 1202 bp mRNA PLN 12-MAY-1997  
DEFINITION P. trichocarpa mRNA for anionic peroxidase Pxp22.  
ACCESSION X97349  
VERSION X97349.1 GI:1279649  
KEYWORDS peroxidase.  
SOURCE Populus balsamifera subsp. trichocarpa.  
ORGANISM Populus balsamifera subsp. trichocarpa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Violales; Salicaceae; Populus.  
REFERENCE 1 (bases 1 to 1202)  
Christensen, J.H., Bauw, G., Boerjan, W. and Van Montagu, M.  
Purification and cloning of poplar xylem peroxidases  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 1202)  
Christensen, J.H.  
TITLE Direct Submission  
JOURNAL Submitted (19-Apr-1996) J.H. Christensen, University of Gent,  
Laboratory of Genetics, Ledeganckstraat 35, Gent, B9000, Belgium  
FEATURES  
Location/Qualifiers  
1..1202  
/organism="Populus balsamifera subsp. trichocarpa"  
/sub\_species="trichobol"  
/db\_xref="taxon:3694"  
/clone="Pxp22"  
/issue\_type="xylem"  
10..1041  
/EC\_number="1.11.1.7"  
/function="signal for ER"  
/codon\_start=-1  
/product="peroxidase"  
/protein\_id="CAA66035.1"  
/db\_xref="GI:1279650"  
/db\_xref="SPTREMBL:Q43100"  
/translation="MHISKAIVEAFPEVVLRLRGTYLACGOLPTPEYDQCTPNVNSIIRD  
VVTETLVSDPRIGASLIRLHFHDCPVNGDGLLDNDPTVISEKEAGNNNSARFE  
VVDTKALLESACPAIVSCADITLIAEESVVLGAPNPTVILGRDSTYASRDANA  
FLAPPEFTLDQRESFTNLSNNSDVLVSGAHFGRKCSFTPERLYDENSTGAPD  
PSIDTTLIAALOELCPGNGSVITDIDISTPDADSDVYSNLDGNRGLIQDDELFS  
TPGADVIALVVAFAFNQTAPESEFVESIRMGNTLSPLTGTGEBIRLNCNVNANLAG  
PDSMLVSSI"

BASE COUNT 348 a 240 c 269 g 345 t  
ORIGIN

Query Match 32.8%; Score 304.4; DB 7; Length 1202;  
Best Local Similarity 58.1%; Pred. No. 1.8e-81;  
Matches 536; Conservative 0; Mismatches 386; Indels 0; Gaps 0;

QY 4 caattaacccctatcttcaacaatagctgcccaagctgcacaacatcgttcgcgac 63  
Db 82 CACCTACTCTCAACATTTTATGACACGACATGTCCAAATGTGACGACATTAATCCGGAT 141  
QY 64 acaatcgtaacagctcagatccgatacccaagatcgctgtcttaatatagctgcac 123  
Db 142 GTCATTACGAGACATTGTGATCCGATCTCGGATTGGAGCCGCTCATCAGGCTCCAC 201

QY 124 ttccatgaactgtctcgtgaatggctgcgaagctagcaatattacttgacaacaccacag 183  
Db 202 TTCCATGACACTGCTTTGTATATGCGTGTGATGATGATGATGATGATGATGATGATGAT 261  
QY 184 ttccgaactgaagaagatgcatctcgggaacgcctaacagcgccaggggcttccagtgatc 243  
Db 262 ATAGTAGGCGAAGGAAGGAGCGGAGGGAACACACTCTGCAAGAGGTTTGAAGTTGTT 321  
QY 244 gatcgtatgaagctgcgcgttgatgtagcaatgcgcccaagcaagctgctgtgagacctg 303  
Db 322 GATACATAGGAAGGCTTTGTTGGAGAGTGCCTGCTGCTACTGCTTTCCTGCTGCTGATA 381  
QY 304 ctgcatgaagctgcgcaacagcagcgtgactctcgcagggcgaccgctccctgagtgccg 363  
Db 382 CTCACATATTCACCTGACATGACATGATCTGTTTGGCAGAGAGCTCAAAATTTGACAGTTCCA 441  
QY 364 ctgcgtcgaagtgactccctcctacagagcatctcctagatctggtcccaacgcctgacctgct 423  
Db 442 TTAGGAAGAAGATGAGACACACAGCAGCGAGATGACAAATGCTTTCCTCCAGCC 501  
QY 424 ccaattcaccctcgcgccagctgaagatagctttaagaacgtggtgctgtaactgcctcg 483  
Db 502 CCTTTTTCACCTTGATCAACTCAGAGAGAGCTTCACTAATGTCAAGCCTTAATAATAT 561  
QY 484 agtgaactgtgctcgtctgctcggagagacacacattggaaagacagctgtaagttcatc 543  
Db 562 AGTGATTTGTACTCTCATCTGTGCTGCACACATTTGGGGGCGCAAAATGTTCTACATTC 621  
QY 544 atgatatagctctacacatlttcaagaacactgggttlaacttgaaccacgcgtgaacactag 603  
Db 622 GACTTCGATTTGATATATTTCAACAGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 681  
QY 604 tatctccgaacactgaagagctgtgtgcccaactgaatggcaacctcagctgactagtgagc 663  
Db 682 CTCTTAGCACACTTTCAGGATATATGCTCCGGAAGGTGGAATGGAGTGTGATACAGAT 741  
QY 664 ttgatctcggagaccccaacactctcgtatgaacaaatgactatgtgaactgaaggagcag 723  
Db 742 CTTGATCTTCACACACCTGATGCTGCTTTGACAGTGTACTACTCTCAATGCTGCAAGGTAC 801  
QY 724 aaagcctgatacagaagtgatacagaactgtttagcagtlcacaagcccaactgaacacatc 783  
Db 802 CGAGGCGCTGCTTCAGACTGATCAAGAACTGTTTCAACTCGGGGGGAGATGATGTGCTAT 861  
QY 784 ccaatggtgagaagtttcttaactactcaaaccttcttaagcgttcgtggaagcc 843  
Db 862 GCACCTGTTTAACGCTTTCAGTCTTAATCAAAACAGCTTTCTTGAAAGCTTTGTGAGATCC 921  
QY 844 atggaccgtctggtgatacatattaccctctgaagcgtgtaaccaagccaagatctgctgaac 903  
Db 922 ATGATAAGATGGGAATGTCACGCTTCGACAGAACTGAAGAGGAGATCAATGATTCAC 981  
QY 904 tgcagagtggtcaacagcaact 925  
Db 982 TGCAGCGTACTCAATGCAAACT 1003

RESULT 14  
POPP01  
LOCUS POPP01 1143 bp mRNA PLN 06-FEB-1999  
DEFINITION Populus nigra mRNA for peroxidase, complete cds.  
ACCESSION D83224  
VERSION D83224.1 GI:1199775  
KEYWORDS peroxidase.  
SOURCE Populus nigra (variety: Italica Koehne) cDNA to mRNA.  
ORGANISM Populus nigra  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Violales; Salicaceae; Populus.  
REFERENCE 1 (bases 1 to 1143)  
Sakuma, Y.  
TITLE Direct Submission



Query Match 32.1%; Score 297.2; DB 7; Length 1371;  
Best Local Similarity 58.5%; Pred. No. 2.8e-79;  
Matches 537; Conservative 0; Mismatches 378; Indels 3; Gaps 1;

Job time: 13146 sec

```
OY 4 caagtaaaccctacatctcagacaatagctgtccaaagctgtccaaacatcgttcgcgac 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 CAGCTGACTCCACATCTACAGATGAAATGCCGAAATGAGCACCACATCATTCGAGGA 197
OY 64 acaatgctcaagagcctcaatccgatcccaagagtcgtcttcaatatacgtctgcac 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198 GPCCTTGCTCAGGCTTTGCAAGACGATCCAGATTCGAGCCGACCTCATTCGACTTCAT 257
OY 124 ttccatgacgtctcgtgaatgtgtgcagcgtacgatalatctgagacaacaccacagt 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 TTCCATGATTTGTTTGTGATGTTGATGATGATCAATTTCTTTGGACAACTGATACT 317
OY 184 ttcgcacgtgaaaagatgcatcggagacgtacacagcgcaggggcttccagtatc 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 318 ATAGAGAGTGAAGAAAGAGCTGCCAAATATATATTCAGCAAGGGGTTTGTATGTTGT 377
OY 244 gatcgcatgaaggtcgtcgttgatgcagcatgcccagaaagtcagltgtgcagacctg 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 GATTAACATGAAGGCTGCAGTAGAGAAATCTTGCCCTGGATAGTTCTGTGATATTT 437
OY 304 ctgactatactgcgcacacagagcgtgactcctgtgcagcgagccgtccttgaaagtgccg 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 CTCGCCATTTGCAGCTGAAGAGTCTGTTGCTTTGGCAGAGAGACCCCTCATGACAGTTCT 497
OY 364 ctgcgtcagcgtgacccctcctacagcattccctagatctggcacaagccttgctgct 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 498 TTGGGAGAGAGAGACAGCTTAATAGCAAAACAGAGTGCTGCTAATCTTCATATCCGGCT 557
OY 424 ccaattctcaaccctgcccacagctgaaagatagctttagaaaagctgggtctgaatcgctcg 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 CCCCTCGAGAGCCTTGGCTGCTCAAAATCGAAAGTTTGACGCCGTGGCCCTCAACACCACT 617
OY 484 agtgaacctgtgctcgtcgcgagagacacacatttggaaaagaccagtgtagttcattc 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 618 AGTACCTGTGTCGCTTTCAGTGCTCACACATTTGGAAGAGCTCAATGTTTAATTTTC 677
OY 544 atgataggctctcaaatctcaagcaacactggttaccctgaccccaagctgaaactacag 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 678 ATCAGTAGATTGTATATTTAGTGCAAGTGCAATCCTGATCCACATTTGAACACAAACC 737
OY 604 tatctcaagacactgaaagctgtgtcccactgaatgcaacctcaatgcaactagtgagac 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 738 TATCTAGCACACACTTCAACAATTAATGCTCAAGGTGAAATCGAGGTGCTTAAACAAC 797
OY 664 ttgatctgagaccccaacacatcttcgataaacaagfacatgtgaatctagaagagcag 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 798 CTTGATGCGACTACACCGATACTTTCGATGGCAACTATTTTCCAAATCTTCAACCAAT 857
OY 724 aaagcctgatacagatgtaacaaagactgttagcagttccagaagcgcactgacacatc 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 858 GAAGGCTGCTTCAGAGTGATCAAGAGTGTCTCTACTACAGGTGCC--GATACCATTT 914
OY 784 ccaetggtgagagtttgcctaactcaactcaaaccttcttlaagccttcogtgaagcc 843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 915 GCCATTTGTCAACATTTTACTAGTAATCAGACTGCTTTCTTGAAGCTTTGTGTGTC 974
OY 844 atgagccgtatggtlaacatlaaccctctgaaggttacccaagagccagatcgtctgaac 903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 975 ATGATTAGAAATGGGAAATATTAGCCCTTAACCTGGACAGTAGAGATCAGATTGAAT 1034
OY 904 tgcagagtggtcaacagc 921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1035 TGCAGAAATAGTCAACAAAC 1052
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 06:40:28 ; Search time 101.52 Seconds

(without alignments)  
2284.555 Million cell updates/sec

Title: US-09-246-451-16

Perfect score: 927

Sequence: 1 atgcagtaaccctacatc.....gagtgtcaacagcaactct 927

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	925.4	99.8	955	1 N90760	Synthetic horserad
2	922.4	99.5	1055	1 N90762	Synthetic horserad
3	843.8	91.0	1072	1 Q37614	Encodes HRP/IGF-II
4	563	60.7	927	1 N93078	Sequence encoding
5	525.6	56.7	924	1 N81508	Horseshadish peroxi
6	485	52.3	501	1 N90761	5' end half of syn
7	446.4	48.2	474	1 N94368	3' end half of syn
8	239.8	25.9	1244	1 V30460	Soybean seed coat
9	239.8	25.9	1244	1 V59077	Soybean seed coat
10	230.4	24.9	1355	1 V63720	Maize per5 root pr
11	163.4	17.6	1270	1 087101	The Tpa gene. Tran
12	163.4	17.6	1270	1 T63685	Full-length Tpa-1
13	163.4	17.6	1270	1 X03970	Full length anionl
14	161	17.4	1348	1 Q39021	Rice green leaf pe
15	159.4	17.2	388	1 Q37615	Encodes HRP/IGF-II
16	148.6	16.0	1131	1 V81611	Basic cucumber per
17	148.6	15.9	1131	1 V62837	Cucumber peroxidase
18	147.4	15.9	1131	1 006207	Cucumber peroxidase
19	146	15.7	1181	1 V81616	Cucumber peroxidase
20	146	15.7	1181	1 V62867	Cucumber peroxidase
21	129	13.9	1064	1 087093	Tpa-1 phase insert
22	129	13.9	1064	1 T63677	Peroxidase gene. T
23	129	13.9	1064	1 X03962	Antionic tobacco pe
24	119.8	12.9	1326	1 T66685	Soybean peroxidase
25	119.8	12.9	1326	1 V69274	Soybean peroxidase
26	119.8	12.9	1326	1 V81421	Soybean peroxidase
27	118	12.7	3359	1 V59078	Soybean seed coat
28	118	12.7	4700	1 V30461	Soybean seed coat
29	114.4	12.3	1144	1 T95782	Soybean peroxidase
30	107	11.5	1167	1 T66687	Soybean peroxidase
31	107	11.5	1167	1 V69276	Soybean peroxidase
32	107	11.5	1167	1 V81423	Soybean peroxidase
33	105.4	11.4	1315	1 T66684	Soybean peroxidase

34	105.4	11.4	1315	1 V69273	Soybean peroxidase
35	105.4	11.4	1315	1 V81420	Soybean peroxidase
36	98.2	10.6	1191	1 T66686	Soybean peroxidase
37	98.2	10.6	1191	1 V69275	Soybean peroxidase
38	98.2	10.6	1191	1 V81422	Soybean peroxidase
39	88	9.5	6550	1 V63717	Maize per5 root pr
40	79.8	8.6	686	1 V81617	Cucumber peroxidase
41	79.8	8.6	686	1 V62868	Cucumber peroxidase
42	75.2	8.1	278	1 V23835	Plant POX enzyme D
43	71.4	7.7	687	1 V23906	Plant POX enzyme D
44	60	6.5	688	1 V23907	Plant POX enzyme D
45	59.6	6.4	1410	1 V23905	Plant POX enzyme D

## ALIGNMENTS

RESULT	ID	Score	Query Match	Length	Description
1	N90760	99.8%	925.4	955	Synthetic horserad
2	N90760	99.5%	922.4	1055	Synthetic horserad
3	N90760	91.0%	843.8	1072	Encodes HRP/IGF-II
4	N90760	60.7%	563	927	Sequence encoding
5	N90760	56.7%	525.6	924	Horseshadish peroxi
6	N90760	52.3%	485	501	5' end half of syn
7	N90760	48.2%	446.4	474	3' end half of syn
8	N90760	25.9%	239.8	1244	Soybean seed coat
9	N90760	25.9%	239.8	1244	Soybean seed coat
10	N90760	24.9%	230.4	1355	Maize per5 root pr
11	N90760	17.6%	163.4	1270	The Tpa gene. Tran
12	N90760	17.6%	163.4	1270	Full-length Tpa-1
13	N90760	17.6%	163.4	1270	Full length anionl
14	N90760	17.4%	161	1348	Rice green leaf pe
15	N90760	17.2%	159.4	388	Encodes HRP/IGF-II
16	N90760	16.0%	148.6	1131	Basic cucumber per
17	N90760	15.9%	148.6	1131	Cucumber peroxidase
18	N90760	15.9%	147.4	1131	Cucumber peroxidase
19	N90760	15.7%	146	1181	Cucumber peroxidase
20	N90760	15.7%	146	1181	Cucumber peroxidase
21	N90760	13.9%	129	1064	Tpa-1 phase insert
22	N90760	13.9%	129	1064	Peroxidase gene. T
23	N90760	13.9%	129	1064	Antionic tobacco pe
24	N90760	12.9%	119.8	1326	Soybean peroxidase
25	N90760	12.9%	119.8	1326	Soybean peroxidase
26	N90760	12.9%	119.8	1326	Soybean peroxidase
27	N90760	12.7%	118	3359	Soybean seed coat
28	N90760	12.7%	118	4700	Soybean seed coat
29	N90760	12.3%	114.4	1144	Soybean peroxidase
30	N90760	11.5%	107	1167	Soybean peroxidase
31	N90760	11.5%	107	1167	Soybean peroxidase
32	N90760	11.5%	107	1167	Soybean peroxidase
33	N90760	11.4%	105.4	1315	Soybean peroxidase

```
Db 191 AGTTTCGGCACTGAAAGATGCAATTCGGGAGCCGTAACAGCCGACGGGCTTTCAGTg 250
Oy 241 atcgatcgatgaagctgcgcttgaagtcgaagatgccacgaacagctgctgtgcagac 300
Db 251 ATGCATGCAATGAAGGCTGCCGTTGAGTCAAGCATGCCACAGACATGATGTGCGAGAC 310
Oy 301 ctgctgactaagctgcgcaacagaagcgtgactctgtcagcgagaccgtccctggaagtg 360
Db 311 CTGCTGACTATAGCTGCGCAACAGAGCGTGAATCTTCAGCGGAGACGCTCCTGGAGAGTG 370
Oy 361 ccgctgctgcaagctgactccctcaaggaacttctctagaatctgagcaacgccaacttgc 420
Db 371 CCGCTCGGTGAGCTGAGCTCCCTACAGGCAATCTCTAGATCTGGCGCAACGCCAATCTTCT 430
Oy 421 gctccattctcaaccctgcccacagctgaagatagctttagaagacgtgggtctcaatcgc 480
Db 431 GCTTCATTTCTTACCCCTGCCAGCTAAGATAGCTTTGAAACGTGGGTCTTAATCGC 490
Oy 481 tcgaatgacctgtgtgctctgtccgaggaacacacatttggaaagaaccagtgtaggttc 540
Db 491 TCGAGTGAACCTTGTGGCTGTGTCGCGAGAGACACATTTGGAAAGAACCAAGTGTAGGTTTC 550
Oy 541 atcatgataagctctcaaatctcagaacacactgggttacctgaccccaagctgaaact 600
Db 551 ATCATGGATAGGCTCTAACATTTCTAGCAACACTGGGTTACCTGACCCACGCTGAACACT 610
Oy 601 acgatctcagaacactgaagagcctgtgtcccaactgaatggaacacctgaatgacactgtg 660
Db 611 ACGATGTCAGACACAGAAAGGCTTGTGCCACTGAAATGGCAACCTCAATGCGCATAGTG 670
Oy 661 gaacttgatctgaggaacccaacacatctcgaatacaagtaactatgtgaactagaagag 720
Db 671 GACTTTGATCTGGGAGCCCAACCATCTTGATCAAGTACTATGTAATCTAGAGAGAG 730
Oy 721 caaagaagcctgtatagagtgatgaagaactgttagcagctcagaagcgcacatgaaacc 780
Db 731 CAGAAAGCCTGATACAGATGATCAAGAACTGTTTAGCAGTCCAAAGCCACCTGAGACCC 790
Oy 781 atcccaatgctgaagaatttctgaactctactcaaacctctttaaagccttcgtgaa 840
Db 791 ATCCCATGCTGAGAAATTTTGGCAACTCTACTACCAACTCTTTTAACGGCTTGTGTGAA 850
Oy 841 gccatggaacgtatggtaacataaccctctgacgggtacccaagccagatctgctgt 900
Db 851 GCCATGGAACGCTATGGGTATCAATTACCCCTCTGACGGGTACCCAGGCGCAGATTCGTCTG 910
Oy 901 aactgcagatgtgaacagaactct 927
Db 911 AACTGCAGATGTGTAACAGCAACTCT 937

RESULT 2
N90762
ID N90762 standard; DNA; 1055 BP.
AC N90762;
DE 18-JUL-1990 (first entry)
DE Synthetic horseradish peroxidase gene (HRP) modified for expression
DE in mammalian cells
KW horseradish peroxidase; restriction sites; DNA construct;
KW fusion protein; enzyme assays; immunoassays; synthetic gene;
KW synthetic linker; PSD18.
OS horseradish.
FH key
FT misc_feature
FT 1..73 Location/Qualifiers
FT /*tag= a
FT /*note="linker used to adapt synthetic gene"
FT misc_feature
FT 933..1055
FT /*tag= b
FT /*note="as above"
FT cds
FT 12..68
FT /*tag= c
FT /*product=N-terminal pre sequence
FT 933..1040
FT cds
```

```
FT /*tag= d
FT /*note="C-terminal pro sequence"
FT mat_peptide
FT 69..992
FT /*tag= e
FT /*note="claimed in Claim 2"

WO8903424-A.
PD 20-APR-1989.
PF 07-OCT-1988; G00833.
PR 08-OCT-1987; GB-023662.
PA (BRB1-) Brit Bio-Tech Ltd.
PI Edwards RM, Burke JF;
DR WPI; 89-130042/17.
DR P-PSDB; P95203.
PT Synthetic gene encoding horseradish peroxidase -
PT used for efficient produ. of the enzyme or fusion prods. for
PT use in enzyme assays, esp. immunoassays
PS Disclosure; ; 45pp; English.
CC The synthetic gene has many restriction sites. The synthetic gene is
CC prep'd. by coupling successive nucleotides and/or ligating
CC appropriate oligomers. The front end of the synthetic HRP gene was
CC modified by the replacement of the HindIII-HpaI fragment (see n90760)
CC with a synthetic linker carrying an NdeI site on the initiator AUG. The
CC synthetic HRP gene was cloned into the expression vector p6C517 on a
CC NdeI-BamHI fragment to give the plasmid PSD18. The plasmid was maintained
CC E.coli strain W3110 lacIq and high level of expression of active HRP
CC was obt'd. A construct containing the claimed sequence (See FT) is
CC claimed. The construct can be a vector where the HRP gene and a
CC desired gene are co-expressed as a single fusion prod.
CC a single polycistronic message, or 2 separate but linked transcriptional
CC units. The fusion proteins are for use in enzyme assays, esp. immuno-
CC assays. Genes carrying mutations that destroy HRP activity can be used
CC to follow the frequency of reversion or suppression of the particular
CC mutation introduced into the gene. The constructs can be used in
CC transgenic animals, esp. mammals.
SQ Sequence 1055 BP; 260 A; 294 C; 250 G; 251 T;

Query Match 99.5%; Score 922.4; DB 1; Length 1055;
Best Local Similarity 99.9%; Pred. No. 5.5e-290;
Matches 923; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 cagttaaccctaatctctacgaacatagctgtcccaagctgtccaacatcgttcgagac 63
Db 69 CAGTTAACCCCTTACATCTACGACAAATAGCTGTCCAACTGTCCAAATCTGTCCGAC 128
Oy 64 acaatcgtaacagagctcaaatcgcgatccagatgcgtcttcaataatgcttgacac 123
Db 129 ACAATCGTCAACGAGCTCAAGATCCGATCCAGATCCGCTCTTCAATATTACGTTCAC 188
Oy 124 ttccatgactgcttcgtgaatgtgttgacgctagcatatctactggaacaacacagat 183
Db 189 TTCCATGACTGCTTCTGGAATGTTGGAGCGTATAGCATTTACTGGAACACACCACAGT 248
Oy 184 ttccgaactgtaaaagatgcatctcgggaacgcgtacacagcgccaggggtcttcagtgatc 243
Db 249 TTCCCGCACTGAAGAGATGCAATTCGGGAAACGCTTAACAGCCGAGGCGCTTCCAGTATGC 308
Oy 244 gatcgaatgaagctgctgctgttgatgaacatgcccacgaacaaatcagttgttgagactg 303
Db 309 GATCGCAATGAAGGCTGCGCTTGAGTCAAGCCTCCACAGAACACTGATTTGTGAGACCTTG 368
Oy 304 ctgactaagctgcgcaacagagcgtgactctgtcagcgagcgacgctccttgaaagtgcg 363
Db 369 CTGACTATAGCTGCGCAACAGAGCGTGAATCTTGCAGGCGGACCGTCTGTGAAGTGGCCG 428
Oy 364 ctcgatcgacgtgactccctcacaagcaatctcagatctggtccaaagcgaactgtgctgt 423
Db 429 CTCGCTGAGAGTGAATCCCTACAGGCAATCTCTAGATCTGGCAACGCCCAACTTGCTGCT 488
Oy 424 ccaatctcaacccctgcccagctggaagatagctttagaagatgggtctgaaatcgtctg 483
Db 489 CCAATCTTACCCCTGCCCCGAGGAGGATAGCTTTAGAAACGCTGGGTGATGATGCTGCTGC 548
```



QY 484 atgagacctgtgctcgtcccgagagacacacattggaaagaacagtgtagtcatc 543  
|||||  
Db 549 AGTCACCTTTGGCTCTGCTCCGAGACACACATTGGAAAGAACCGTGTAGCTTCATC  
544 atgatagagctcacaatttcagacaacatgggttacctaaccccaagcgaacacagc 603  
|||||  
Db 609 ATGGATAGGCTCTACATTTCAGACACACTGGGTTACTCTGACCCACGCTGAACTACG 668  
604 tatctccagacactgtagagagctgtgcccactgaaatgcaacctcagtgcactagtgac 663  
|||||  
Db 669 TATCTCCAGACACTGAGAGAGCTTGTCGCCCACTGAAATGGCAACCTCATCTAGTGTGAC 728  
664 ttgatctgcgagaccaccaacatcttcgataaacaagtaactatgtgaactagaggagcag 723  
|||||  
Db 729 TTTCATCTGGGAGCCCAACCATCTTGTGATACAAAGTACTATGTGAATCTAGAGAGCAG 788  
724 aaaggacctgataagagtgatgataagaactgttagcagctcagagccactgaaacatc 783  
|||||  
Db 789 AAAGGCTGTATACGAGATGTATCAAGAACTGTTAGCAGTCCAAACGCCACTGACACGATC 848  
784 ccactgtgagaaagtttctgaactctcactcaaacctctcttaacgcctcgttgagagcc 843  
|||||  
Db 849 CCACGTGTGAGAACTTTTGTCTACTACTCAAACTTTCTTTAACGCCCTTGTGGAAGCC 908  
844 atgagccgataggtaacattaccctctgacggttacccaaggccagattcgtctgac 903  
|||||  
Db 909 ATGGACCGTATGGGTATACATTACCCCTCTGTACGGGTACCCAGGCCAGATTCTGTCGAC 968  
904 tgcagagtggtcaacagcaactct 927  
|||||  
Db 969 TGCAGAGTGGTCAACGCAACTCT 992  
|||||

RESULT 3  
Q37614  
ID Q37614 standard. DNA: 1072 BP.  
AC Q37614;  
DT 23-JUN-1993 (first entry)  
DE Encodes HRP/IGF-II fusion protein.  
KW Insulin receptor; E.coli periplasmic protease resistance;  
KW wound healing; osteoporosis; muscle-wasting; growth disorder;  
KW horseshadish peroxidase; inclusion bodies; ds.  
OS Synthetic.  
FH Key  
FT Location/Qualifiers  
FT 11..1060  
FT /\*tag= a  
FT /note= "HRP/IGF-II fusion gene"  
FT misc\_feature  
FT 11..856  
FT /\*tag= b  
FT /function= HRP coding sequence  
FT /note= "encodes first 282 amino acids of HRP"  
FT misc\_feature  
FT 857..1060  
FT /\*tag= c  
FT /function= IGF-II coding sequence

FT WO9303152-A.  
PN 18-FEB-1993.  
PD 27-JUL-1992; G01389.  
PF 29-JUL-1991; GB-016325.  
PR 05-FEB-1992; GB-002401.  
PA (BR1-) BRITISH BIO-TECHNOLOGY LTD.  
PI Bawden L, Edwards RM;  
PI MPI: 93-076509/09.  
DR P-PSDB: R32760.  
PT New insulin-like Growth Factor II analogues - for treating growth  
PT disorders, osteoporosis and muscle wasting diseases; have less  
PT affinity for insulin receptors than unmodified IGF-II  
PS Example 4: Page 44-46; 71pp; English.  
CC This sequence encodes a fusion protein comprising the first 282  
CC amino acids of HRP fused to IGF-II via a methionine residue to  
CC allow for the release of mature IGF-II by cleavage with CNBR. The  
CC fusion lacks the final 27 amino acids of HRP. This fusion protein  
CC was not ideal for production of IGF-II; the IGF-II is only 19% of  
CC the total fusion protein and the HRP sequence contains two internal

CC methionine residues which act as additional cleavage sites.  
CC The IGF-II is the wild-type sequence, i.e. with Arg residues at  
CC positions 37 and 38. The invention covers analogues of this sequence  
CC in which one or both of the Arg residues are subd. by other amino  
CC acids (pref. non-basic and esp. Gln). The mutants are protease-  
CC resistant and have reduced affinity for insulin receptors.  
SQ Sequence 1072 BP; 259 A; 304 C; 255 G; 254 T;

Query Match 91.0%; Score 843.8; DB 1; Length 1072;  
Best Local Similarity 99.8%; Pred. No. 2e-264; Indels 0; Gaps 0;  
Matches 845; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgcagttaacccccctacatctcacaataagctcgtcccaagtgccaacatcgttcgc 60  
|||||  
Db 11 ATGCAGTTAACCCCTTACTATTCTACGACATATAGCTGTCCCAACGTCATCATCTGTCGC 70  
61 gacacaatcgtcaagagctcgaatccgatcccaaggatcgctgtcaataatactgtcgtc 120  
|||||  
Db 71 GACCAATCGTCAACGAGCTCAGATCCGATCCAGATCCGCTCTCATATTATGATGCTGTG 130  
121 cacttcgatgactgtctcgtgaatggttgagagctgagctgatatattactggacaacacc 180  
|||||  
QY 131 CACTTCATGACTGCTTGTGAAATGTTGCGACGCTAGCATATTATGAGCAACACCACC 190  
|||||  
Db 181 agttccgcaactgaaagagatgcatcgggagcgctaacagcgcgagggccttcacagt 240  
|||||  
QY 191 AGTTTCCGCACTGAAAGATGATGCTTGGGAGACCTTAACACGCCGACGGCTTTCAGATG 250  
|||||  
QY 241 atcgatcgatgaagagctgcggttgatgcatcagcatcgtcccaagacagtcagttgtcagac 300  
|||||  
Db 251 ATCGATCGCATGAAGGCTGCGGTTGAGTCAAGCATGCCCCAGCAACAGTATGTTGACGAC 310  
301 ctgctgactatagctgctgcgaacagagcgtaactctgcagggcggagccgtctgagagtg 360  
|||||  
Db 311 CTGCTGACTATAGCTGTGCGCAACAGAGCGTGACTCTTTCAGCGGACCGTCTGTGAGAGTG 370  
361 ccgctgctgcaagctgactccctcctacaggaatcctctagatctcggccaagcccaacttgc 420  
|||||  
QY 371 CCGCTCGGTGCACGTGACTGCTCCCTACAGGACTCTTAATCTGGCCACGCCCACTTGTGCT 430  
421 gctccatctctcacccctgcgcccagctgaagatagctcttgaagaacgtggtgtcgaatcgc 480  
|||||  
Db 431 GCCTCATTTCTTACCCCTGCCCCAGCTGAAGATAGCTTTGAAACGTTGGGTGAATCGC 490  
481 tgcagtgacctgtgtgctcgtcgcgagagacacacattggaaagaaccagtgtagttc 540  
|||||  
QY 491 TCGAGTGACCTTGTGGCTGTGTCGAGGACACACATTTGGAAAGAACCAAGTTAGGTTTC 550  
541 atcaatgataaggtctcacaatttcagaacaacatggttacctgaccccaagctgaacact 600  
|||||  
Db 551 ATCATGATATAGGCTCTACAAATTTAGCAACACTGGTGTACTGTGACCCCAAGCTTAACACT 610  
601 acgatctccagacactgagagagctgtgcccactgaaatgccaacctcagtgactagtg 660  
|||||  
QY 611 ACGTATCTCCAGACACTGAGAGGCTTGTGCCCATGTGATGGCAACCTCAGCTAGTG 670  
661 gacttgatctgcgagcccaacacatcttcgataaacaagtaactatgtgaactagaggag 720  
|||||  
Db 671 GACTTGTATCTGCGACCCCAACCATCTTGTGATTAACAGTACTATGTGAATCTAGAGAG 730  
721 cagaaagggcctgataagagatgataagaactgtttgagcagtcagagccagcagaaacc 780  
|||||  
QY 731 CAGAAAGGCTGTATACGAGATGATCAAGAACTGTTTGGCAAGCCCAAGCTGACACC 790  
781 atcccaatgtagaagcttctgtaactctactcaaaactcttctaaagccttcgttgaa 840  
|||||  
Db 791 ATCCCATGTGGGAAGAAATTTTGTGTAACCTACTAATAAACCTTCTTAACGCCCTTGTTGAA 850  
841 gccatgg 847  
|||||  
Db 851 GCTATGG 857

```

RESULT 4
ID N93078 standard; DNA; 927 BP.
AC N93078:
DT 10-MAR-1993 (revised)
DT 25-JUN-1989 (first entry)
DE Sequence encoding horseradish peroxidase (HRP) isoenzyme C.
KM Horseradish peroxidase (HRP) isoenzyme C; detection reagent;
KW superoxide radicals (SOD); ss.
FH Key
FT cds
   1..924
   location/Qualifiers
   /tag= a
   BP-299682-A.
   PD 18-JAN-1989.
   PF 07-JUL-1988; 306222.
   PR 17-JUL-1987; GB-016938.
   PR (AMER-) Amersham Int. PLC.
   PI Chiswell DJ, Ortlepp SA;
   DR MPI; 89-017184/03.
   P-PSDB: P94810.
PT New synthetic DNA encoding horseradish peroxidase C isoenzyme or variants
PT or active fragments, useful eg as detection reagent.
PS Disclosure; fig.6; 19pp; English.
CC Horseradish peroxidase Isoenzyme C (HRP-C) is produced from this sequence
CC by recombinant techniques. Conservative variant can be encoded by intro-
CC ducing base changes resulting in amino acid (AA) substitutions. Proper-
CC ties of this enzyme are essentially the same as native HRP-C. A peptide
CC conty. at least 1 HRP-C epitope can also be encoded by restriction enzyme
CC cleavage resulting in peptides of different lengths (conty. active AA's
CC Arg-183 and Tyr-185). HRP-C is a scavenger of toxic superoxide radicals
CC (SOD) and is useful as a detection reagent. It provides enhanced illumina-
CC tion by catalysis of a substrate such as luminol, in the presence of a
CC light-output enhancer.
SQ Sequence 927 BP; 247 A; 262 C; 199 G; 219 T;

```

```

Query Match 60.7%; Score 563; DB 1; Length 927;
Best Local Similarity 75.6%; Pred. No. 3.4e-173;
Matches 698; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

```

```

QY 4 cagttaccctacatctctcgaacatagctgtcccaagctgtccaacatcgctgcgac 63
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 CAGTTAACGCCGCGATTCATGATACAGCTGTCCAAACGCTTCGATATTTGTCGGCAT 60
QY 64 acatcgtcaacagctcagatccgatcccaagatcgctctcaatattacgtctgcac 123
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 ACCATGCTGATAGCTGCGGTCGCATCCACGATTTGACAGTACATCTTTCGTTACAT 120
QY 124 ttccatgactgctctgtaattgtctgcagactagacatattacttgacaacacacacagt 183
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 TTTCATGACTGCTTTGTAACGATGATGACGCGTCATCTCTGGACATACACAGC 180
QY 184 ttccgcaactgaaaagatgattcctgcggaaacgctaacacgcccgaaggtcttccagtatc 243
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 TTCCGAATGAGAAAGCGCATTTCCGGAACGCGAATTTCCGCCCGCGCTTTCCAGTCATC 240
QY 244 gatcgataaagaagctccgcttgaagtcagatcagcgaagacagtcagttgtgaagactg 303
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 241 GATCGCATGAAGCTGCGAGTCGAGTCGGCTGCCCCGAGAACTTTTCGTGTGACAGCTG 300
QY 304 ctgactatagctgcgacaacagagctgactctgcagcgcgagacgtctggaagtgacg 363
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 301 CTGACCATCGCACCCACGAGAAATGGTCAACCTTAGCGGGGGGCGCTCATGGCAGTAGCA 360
QY 364 ctgcgtcagactgactccctcctaagagcatctcagatctggccaagccaaactgtcctgct 423
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 361 TTGAGCAGTCGTGATCCCTACAGGCTTTCTGATGATCTACATCAATCTACAGGGCG 420
QY 424 ccatttaccctcgtcccaagctgaaagatagctttagaaaacgtgggtctgaatgctg 483
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 421 CCATTTTACCTCTACCCACACTTAAGGATTCATTTTCGTACCTAGCGCTGAACCGCTCG 480

```

```

QY 484 agtacccttggtcctgtctcgcgaggacacacatttgaaaagaacagtgtaagttcaltc 543
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 481 AGTATCTAGTACTCTTTCTGTGTGGCCACACTTTGGAAAGAACGATGCCGATTCATC 540
QY 544 atggatagctctcaacatttcgaacactgggtttaccctgaccccaagctgaacataag 603
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 541 ATGGACCGATTTGACAACCTTTTCAACACATGACACTACCGGATCCACGCTCAATFAGGAC 600
QY 604 taltccagaactggagagctgtgcccactgaaatggcaactcagtgcaactagtggac 663
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 601 TACTCTACAGACACTCCGGGACTATGCCCCCTTTAAAGGAAACTTAAGCGCGCTAGTCGAC 660
QY 664 ttgatctgcgagcccaacacatcttcgataaacaagtactatgtgaatctagaagagcag 723
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 661 TTTGACCTAGCTAGCGCAACAAATCTTGACAAACAGTACTATGTAATCTAGAGAACAA 720
QY 724 aaagccctgataagaagtgataagaactglttaagcaagltccagaagccactgaaccatc 783
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 721 AAAGGACTTATTCAAATCTGTACAGAGCTCTCTCATCACCCGAATCAACAGACAGATC 780
QY 784 ccactggtgagaagtttgttaactctactcaactctttaaagccttcgtggaagcc 843
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 781 CCATCTAGTGGCTAGCTTTGCGAACTCAACACAACTTTTCAATGCAATTCGTAGAGCC 840
QY 844 atgacacgtatggttaacattaccctctgacggtgtaaccaaggtccagattcgtctgaac 903
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 841 ATGATTCGATGATGGGAATATACAGCCACATCAGCGGTACCAAGTCAGATCAGCTTGAC 900
QY 904 tgcagagtgtgtaacagcaactc 926
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 901 TGGCAGTGTGTAATCAATCTGAACTC 923

```

```

RESULT 5
ID N81508 standard; DNA; 924 BP.
AC N81508;
DT 05-NOV-1990 (first entry)
DE Horseradish peroxidase structural gene.
KM Horseradish peroxidase; Armoracia lapathifolia;
KW Armoracia rusticana; ss.
OS Armoracia sp.
FH Key
FT cds
   1..924
   /tag= a
   /product=horseradish peroxidase
   J63207386-A.
   PD 26-AUG-1988.
   PF 19-FEB-1987; 037623.
   PR 19-FEB-1987; JP-037623.
   PA (TOYM) Toyobo KK.
   DR MPI; 88-282535/40.
   P-PSDB: P81167.
PT New structural gene of horseradish peroxidase -
PT produced under artificially controlled conditions by gene
PT manipulation.
PS Disclosure; P: Japanese.
CC DNA was preferably isolated from tissue cultures of root cells of
CC Armoracia sp. The gene can be inserted into a plasmid for
CC transformation of suitable host cells. The product (HRP) can then
CC be collected from the culture medium.
CC See also N82195.
SQ Sequence 924 BP; 256 A; 224 C; 187 G; 257 T;

```

```

Query Match 56.7%; Score 525.6; DB 1; Length 924;
Best Local Similarity 73.1%; Pred. No. 4.8e-161;
Matches 675; Conservative 0; Mismatches 249; Indels 0; Gaps 0;
QY 4 cagttaccctacatctcgaacatagctgtcccaagctgtccaacatcgctgcgac 63
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 CAGTTAACCCCTACCTTCAGACGAATTCATGCTCAATTAAGTCTTAACATCGACGGAT 60

```

```
OY 64 acaatcgtaacagagctcagatccgataccagagtcgctgtctcaatataatcgtctgcac 123
    || || || || || || || || || || || || || || || || || || || || || ||
DB 61 ACATATGCAATGAGCTATGATAGATACAGCCCTGTATTTGGCCGAGACATCCTTTCGCTTAC 120
OY 124 ttcatatgactctcgtcgtgaatggtctgcagagctatgatatctatcgtgacaacaccagct 183
    || || || || || || || || || || || || || || || || || || || || || ||
DB 121 TTCCACGACCTGCTTTGTTAAAGGTTTGACACATCGATCTTGTATGACACACACACATCA 180
OY 184 ttccgcaactgaaagagatgatactcgtgggaacgctaacagcgccagcggtcttcagatgac 243
    || || || || || || || || || || || || || || || || || || || || || ||
DB 181 TTTTGCAACAGAGAAAGATGCCCTTTGGAACCCAAACCTCGGCAAGAGATTTCCAGTGATTT 240
OY 244 gatacgaatgaagagctgcgctgtgagtcagcatalcccaagcaagtcagctgtgcagagctg 303
    || || || || || || || || || || || || || || || || || || || || || ||
DB 241 GATAGAAATGAAGCCGCGGTGAGAGTGATGATGCCCAAGAACCGTTTCATCGCGCAGATTTTG 300
OY 304 ctgactatagctgcgcacacagagcggtgactcttcgagcgagagccgttcctgtgagagtgcc 363
    || || || || || || || || || || || || || || || || || || || || || ||
DB 301 CTCACCAATGCAAGCTCAACCAATCTGTCACTTTGGCGGAGGCTCCTTCTTGAGAGATTCCT 360
OY 364 ctgagtcagagctgcctccatacagagctatcctagatctgcgcaacgccaacttcgtcgt 423
    || || || || || || || || || || || || || || || || || || || || || ||
DB 361 TTGGGCAAGAGATGATGATTTACAAAGCATTTTGTGATCTTGTATGCAAAATCTTCCAGCT 420
OY 424 ccactctcaccctgcgcacagctgaagagatgactttagaaacgtggtgctgtaacgcgtcgc 483
    || || || || || || || || || || || || || || || || || || || || || ||
DB 421 CCATTTCTTACACATTTCCACCAACTTAAGACAGCTTTAGAAATGTTGGCTTCACCGCTTCT 480
OY 484 agtgaacttgctgtcgtctgcgcgagagacaacacttgcgaagaacagcgctgtagtctac 543
    || || || || || || || || || || || || || || || || || || || || || ||
DB 481 TCTATATCTCGTTGACACTGTCTCGGGGGCCACACATTTGTTAAATAATCAGTGCTGGTTAT 540
OY 544 atgatatgagctcctaactttcagaacacatggttacctgaaccccaagcgctgaacactagc 603
    || || || || || || || || || || || || || || || || || || || || || ||
DB 541 ATGACACAGATTTATCAACTTCAGCAACACCGCTTTACCGGTCCTTACTCTCAACACTACT 600
OY 604 taltccagaacactgaagagcctgtgcccactgaatlgcaaccccaagtcgactagtgagac 663
    || || || || || || || || || || || || || || || || || || || || || ||
DB 601 TATCTCAAACTTTTCGTGGGACTATGTCTCCCTCAATGATATCTAAAGCGCTTTGGTGGAT 660
OY 664 ttgtatctgcgagacccaacacactctcgaataacagtaactatgtgaacttagagagcag 723
    || || || || || || || || || || || || || || || || || || || || || ||
DB 661 TTTTATCTAGCTAGCGCAACGATTTTTCACMAAAATATATGTAATCTCGAAGACCAA 720
OY 724 aaagcctgtatagagtgatcgaagaactgtttagcagtgccagcgccacatgcacatc 783
    || || || || || || || || || || || || || || || || || || || || || ||
DB 721 AAAGCACTTTATCCAAAGCGACCAAGAGATTTGTCTTAGCCCAATGCGCACTGACACAATC 780
OY 784 ccaactgltgaagaagtttctgaactcactcaaaccttctttaaagccttcgttgaagcc 843
    || || || || || || || || || || || || || || || || || || || || || ||
DB 781 CTTTTGGTGAATCATTTTGGTAATAGCACACAAACATTTCTTCAATGCAATTTGTGGAGCG 840
OY 844 atggaacgtataggtaacattacccctcgtgacgggtatcccaagcgagatcgtctgaac 903
    || || || || || || || || || || || || || || || || || || || || || ||
DB 841 ATGGATTTGGATGGGAAACATTTACCTCTTACAGAGAACTCAAGACAGATCAGTTGTAAT 900
OY 904 tgcagagtggtcaacagcagcactc 927
    || || || || || || || || || || || || || || || || || || || || || ||
DB 901 TGTATGGTGTGTAAGCTCCAACTCT 924
```

RESULT 6  
N90761  
ID N90761 standard: DNA; 501 BP.  
AC N90761;  
DT 18-JUL-1990 (first entry)  
DE 5' end half of synthetic horseradish peroxidase gene (HRP)  
KM Horseradish peroxidase; restriction sites; DNA construct;  
KW fusion protein; enzyme assays; immunoassays; synthetic gene.  
OS Horseradish.  
FH Key location/Qualifiers  
FT cds 11..493  
/\*tag- a

```
PN M08903424-A.  
PD 20-APR-1989.  
PE 07-OCT-1988; G00833.  
PR 08-OCT-1987; GB-023662.  
PI (BRB1-) Brit Bio-Tech Ltd.  
PI Edwards RM, Burke JF;  
DR WPI: 89-130042/17.  
DR P-PSDB: P95200.  
PT Synthetic gene encoding horseradish peroxidase -  
PT used for efficient prodn. of the enzyme or fusion prods. for  
PT use in enzyme assays, esp. immunoassays  
PS Disclosure: ;45pp; English.  
CC The 3' half of synthetic HRP is given in n94368; n90760 gives the  
CC full synthetic HRP gene, which contains many restriction sites.  
CC The synthetic gene is prepd. by coupling successive nucleotides  
CC and/or ligating appropriate oligomers. A construct containing bases  
CC 14-937 of the synthetic gene is claimed. The construct can be a  
CC vector where the HRP gene and a desired gene are co-expressed as  
CC a single fusion prod., a single polycistronic message, or 2 separate  
CC but linked transcriptional units. The fusion proteins are for use  
CC in enzyme assays, esp. immunoassays. Genes carrying mutations that  
CC destroy HRP activity can be used to follow the frequency of  
CC reversion or suppression of the particular mutation introduced  
CC into the gene. The constructs can be used in transgenic animals,  
CC esp. mammals.  
SQ Sequence 501 BP; 120 A; 149 C; 119 G; 113 T;
```

Query Match 52.3%; Score 485; DB 1; Length 501;  
Best Local Similarity 100.0%; Pred. No. 5,6e-148;  
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 atgagtgtaacccctatacttcaagacaatagctgtcccaagtgccaacatgctgc 60
    || || || || || || || || || || || || || || || || || || || || || ||
DB 11 ATGAGTTAAACCCCTTAATTTCTAGCAAAATAGCTGTCTCCCAACGTCCCAACATGCTTCGC 70
OY 61 gaccacatcgtaacagagctcagatccgataccagagatcgctgttcaatatactagctg 120
    || || || || || || || || || || || || || || || || || || || || || ||
DB 71 GACCAATCGTCACAGAGCTCAGATCCGATCCAGATCGCTGCTTAATATTTACGCTCG 130
OY 121 caattcagtagctctcgtgaatggttgctgcgcgtctagacatatacttgcgaacacacc 180
    || || || || || || || || || || || || || || || || || || || || || ||
DB 131 CACTTCCATGACTGCTCTGTGAATAGTTGAGCGAGCTGACATATTAATGACACACACC 190
OY 181 agttccgcactgaagaagatgactcgtgggaagcgtacacagcgccagggcttccagt 240
    || || || || || || || || || || || || || || || || || || || || || ||
DB 191 AGTTTCGCACTGAAAGGATGCAATTCGGGAACCTTAACAGCCGACGGCGCTTCCAGTG 250
OY 241 atcgatcgatgaagagctgcgctgtgagtcagatgcacgcagaagtcagtttgcagagc 300
    || || || || || || || || || || || || || || || || || || || || || ||
DB 251 ATCGATATGCAATGAAGGCTGCCGTTGAGTCAACATGCCCCACACAGATCAGTTGTGCAAGC 310
OY 301 ctgctgactatagctgcgaacagagcgttgaactcttcagcgagacgcttcgtgagagt 360
    || || || || || || || || || || || || || || || || || || || || || ||
DB 311 CTGCTGACTATAGCTGCGCAACAGAGCGTGACTCTTGACAGCGGACCGCTCTGAGAGTG 370
OY 361 ccgctcggtcgaagtgactccctacagagcatcctatagatctgcgaagcccaacttgc 420
    || || || || || || || || || || || || || || || || || || || || || ||
DB 371 CCGTTCGCTGACGCTGACTCTCTTACAGGCAATTCCTATAGATCTGGCCAAACGCCAATTCCT 430
OY 421 gctccactctcaccctgcgcacagctgaagagatgactttagaaacgtggttctgaatcgc 480
    || || || || || || || || || || || || || || || || || || || || || ||
DB 431 GCTTCATTTCTTACACCTTGCCCGCACTGAAGATGACTTTTGAAGACGTGGGTCTGAATGCG 490
OY 481 tcgag 485
    || || || || || || || || || || || || || || || || || || || || || ||
DB 491 TCGAG 495
```

RESULT 7  
N94368  
ID N94368 standard: DNA; 474 BP.  
AC N94368;

DB	18-Jul-1990 (first entry)	Location/Qualifiers
DE	3' and half of synthetic horseradish peroxidase gene (HRP)	
KW	Horseradish peroxidase; restriction sites; DNA construct;	
KW	fusion protein; enzyme assays; immunoassays; synthetic gene.	
OS	Horseradish.	
FH	Key	10..937
FM	cds	/*tag= a
FN	W06903424-A.	
PD	20-APR-1989.	
PF	07-OCT-1988; G008033.	
PR	08-OCT-1987; GB-023662.	
PA	(BRRI-) Brit Bio-Tech Ltd.	
PI	Edwards RM, Burke JF;	
DR	WPI: 89-130042/17.	
DR	P-FSDB: p95200.	
PT	Synthetic gene encoding horseradish peroxidase -	
PT	used for efficient produ. of the enzyme or fusion prods. for	
PT	use in enzyme assays, esp. immunoassays	
PS	Disclosure: :45pp; English.	
CC	The 5' half of synthetic HRP is given in n90761; n90760 gives the	
CC	full synthetic HRP gene, which contains many restriction sites.	
CC	The synthetic gene is prep. by coupling successive nucleotides	
CC	and/or ligating appropriate oligomers. A construct containing bases	
CC	14-937 of the synthetic gene is claimed. The construct can be a	
CC	vector where the HRP gene and a desired gene are co-expressed as	
CC	a single fusion prod., a single polycistronic message, or 2 separate	
CC	but linked transcriptional units. The fusion proteins are for use	
CC	in enzyme assays, esp. immunoassays. Genes carrying mutations that	
CC	destroy HRP activity can be used to follow the frequency of	
CC	reversion or suppression of the particular mutation introduced	
CC	into the gene. The constructs can be used in transgenic animals,	
CC	esp. mammals.	
SQ	Sequence 474 BP; 133 A; 124 C; 106 G; 111 T;	
QY	Query Match 48.2%; Score 446.4; DB 1; Length 474;	
	Best Local Similarity 99.8%; Pred. No. 1.9e-135;	
	Matches 447; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
DB	9 CTCGGTACCTTGGCTCTGTCGGAGACACACATTGGAAAGAACACGTGAGTTT	68
QY	440 CTCGGTACCTTGGCTCTGTCGGAGGAGACACATTGGAAAGAACACGTGAGTT	539
DB	540 CATCATGATAGGCTCTCAATTTCAGCAACACTGGTACTGACCCAGCTGAACAC	128
QY	600 taagtattctccagacactgtgagagcttgtgccactgaatggcaacctgaigtacagt	659
DB	129 TACGATATCTCCAGACACTGAGAGGCTTGGCCCACTGATGGCAACCTCAGTCACTAGT	188
QY	660 ggaatttgaatctcgtggagaccaccaacatcttggataaagaactcgtgaaactcagaaga	719
DB	189 GGACTTTGATCTGCGCGAGCCCAACCATCTTCGATATGAACAGTATGTGAAATCTGAGAGA	248
QY	720 gcaagaagcctgatacagaagtgatcaagaactglttagcagtcacagacgcaactgacac	779
DB	249 GCAGAAAGGCCCTGATACAGAGTGAATCAAGAACGTTTGTAGCACTCCAAACGCCACATGACAC	308
QY	780 catccactgtgagaaagtttgcctcaactctcctaacaacctcttaagcgctctgtga	839
DB	309 CATCCCACTGGTGAAGAAAGTTTGTCTACTTACTCAACCTTCTTTAAAGCCTTGATGGA	368
QY	840 agccatggacccglaatggglaaatatcaccctctgacgggtatcccaaggccagatcgct	899
DB	369 AGCCATGGACCCGTATGGGATACATTACCCCTCTGACGGGTACCAAGGCAGATTGCTCT	428
QY	900 gaactgcagagtgtcacaagaactctc 927	
DB	429 GAACGTGACAGAGTGTCAACAGCAACTCT 456	

RESULT	8
ID	V30460 standard; cDNA; 1244 BP.
V30460	
AC	V30460;
DT	09-NOV-1998 (first entry)
DE	Soybean seed coat peroxidase cDNA.
KW	Peroxidase; seed coat; soybean; transgenic plant; ss.
OS	Glycine max cv. Harosoy 63.
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	sig_peptide
FT	1..77
FT	/tag= a
FT	/tag= b
FT	mat_peptide
FT	78..1056
FT	/tag= C
PN	CA2211018-A.
PD	30-MAR-1998.
PE	19-SEP-1997; 211018.
PR	30-SEP-1996; US-723414.
PR	30-SEP-1996; CA-186833.
PA	(MTC ) CANADA MIN AGRICULTURE.
PI	G1jzen M;
DR	WPJ: 98-388720/34.
DR	P-PSDB: W44931.
PT	New DNA used in preparation of soybean seed coat peroxidase - used
PT	e.g. in diagnostic assays, for de-inking waste paper, for oxidation
PT	of primary alcohols, in paper and pulp industries and to encode
PT	herbicid resistance in plants
PS	Claim 1; Page 42-45; 75pp; English.
CC	This cDNA clone includes an open reading frame coding for soybean
CC	seed coat peroxidase. A cDNA library constructed from developing
CC	seed coat tissue of soybean Epp cultivar Harosay 63 was screened
CC	with a probe corresponding to the conserved active site domain of
CC	plant peroxidases to isolate the clone. The structural gene (see
CC	V30461) was isolated from a Harosoy 63 genomic library using the
CC	cDNA as probe. Vectors, host cells, and transgenic plants are
CC	provided, as well as methods for the production of recombinant
CC	seed coat peroxidase. Heterologous genes of interest are produced
CC	within seed coat cells by propagating a transformed plant with a
CC	vector comprising contiguous nucleotides 1-1532 of the genomic DNA.
CC	Soybean coat peroxidases are useful industrially, e.g. their high
CC	thermal stability even at extremes of pH make them useful in
CC	diagnostic assays as an enzyme label for e.g. antigens, and within
CC	staining techniques. They are also useful for de-inking printed
CC	waste paper, biocatalytic oxidation of primary alcohols, as a
CC	replacement for chlorine in the pulp and paper industry or as a
CC	formaldehyde replacement for use in adhesives. The peroxidase may
CC	also be expressed in an organ- or tissue-specific manner within
CC	plants by employing appropriate promoters. Heterologous genes may
CC	encode herbicide resistance, biological control of insects/pathogens
CC	or proteins of commercial interest (e.g. pharmaceuticals) or that
CC	alter e.g. the nutritive value of seeds. The cDNA/genomic DNA are
CC	also useful to produce probes to detect polymorphisms that
CC	distinguish Epp and eep genotypes; varieties containing a
CC	dominant Ep allele accumulate large amounts of peroxidase in the
CC	hourglass cells of the subepidermis, whilst homozygous recessive
CC	eep genotypes do not accumulate peroxidase in the hourglass cells
CC	and have much reduced seed coat peroxidase activity.
SQ	Sequence 1244 BP; 358 A; 261 C; 250 G; 375 T;
Query Match	25.9%; Score 239.8; DB 1; Length 1244;
Best Local Similarity	55.3%; Pred. No. 4.4e-68;
Matches 510; Conservative	0; Mismatches 407; Indels 6; Gaps
OY	4 cagtaaacccctacattactagaacaatagctgtcccaacglttccaacatcttcgcgac 63 
DB	79 CAGCTTACTGCTCCTACGTTCACAGACAAGAATCATGTCCAATATCTGCCATTGGTTTGGA 138 
OY	64 acaatcgtcaacagagctcaagatccgatcccgatgcgtgccttaataatatcagtgtcac 123 
DB	139 GTAATCTTCGATGCATCTTCTTCACCGATCCCCCAATCGGGGCGAGTCATGAGGGCTTCAT 198 

```

QY 124 ttccatgactgctctcgtaaggtcgcagcctagacatactactgagcaaacaccacagt 183
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 199 TTTCATGATGATTGCTTGTCAAGGTGTGATGATGATGATGATGATGATGATGATGAT 258
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 184 ttccgcacgtaaaagatgattcgggaagcgttaacagcgccagggtccttcagtgatc 243
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 259 ATAGAAAGCGAGCAAGATGATGCTTCACAAATATACATTAAGAGATTTGGAGTGTGC 318
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 244 gatcgatgaagcctgcgcttgatcgacgacgtgccacgaacagtcagttgtgcagccctg 303
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 319 AATGACATCAAGACAGCGGTGGGAAATAGTTGTCCAGACACAGTTCCTTGTGCTGATATT 378
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 304 ctgacataagctgcgcaacagagcgtgactctgacgcgagcgagcgtctcgtgagagtgccg 353
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 379 CTTCGATATTGACGTGAATATAGCTTCGTTCTGGAGAGAGCTCCAGAGATGGCCAGTTCCA 438
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 364 ctgcgtgcgaagtgactccctaccagcattccctgatctgagcgaacgcgaactgctgct 423
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 439 TTAGGAAGAAAGGACAGCTTACAGCAAAACCGTTCCAATCAAAACCTTCCAGCA 498
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 424 ccaattctcaacctgcgcacagctgaagatagctttagaacgtgggtgctgaaatcgctcg 483
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 499 CCTTTCTCAACCTCACTCACTTAAAGCTTCCTTGTGCTGTTCAAGGCTTCACAA--CACC 555
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 484 agtgaacctggtgctctgctcgcgagagacacacattggaagaagacagtgtaagttcatc 543
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 556 CTTCGATTTAGTTAGCTTCAGCTGAGTGTGATGATGATGATGATGATGATGATGATGAT 615
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 544 atgataagcctcacaattcagaacaactgggttaacctgagccccaagctgaacaactagc 603
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 616 ATAAACCGATTATACAACTTCACGCAACACTGGAACCTGTATCCAACTCGAACACANACA 675
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 604 taactcagacactgagagagctgtgcccactgaatagtgcaactgaactgaactgaagagac 663
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 676 TACTTGAAATATTGCTGCTGCAAGATGCCCCCAAGATGACACTGGGGCTTAACCTCACCAAT 735
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 664 ttgatctcgcgagcccaacacatctcgtatgaagaactgaactgaactgaactgaactga 723
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 736 TTGGACCTGAGCACACCTGATTCATATTGACACAGATACATTCATTCATTCATTCATTC 795
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 724 aaagcctgatacagagtgatcaagaactggtttacgaagcttcagacgcgcactgaacacatc 783
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 796 AATGGCTTACTTCAGAGTGCACAAAGAACTTTCCTCAGCTCTGCTG--CTGATACCAT 852
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 784 ccaactggtgagaggttttctgtaactctactcaaacctcttgaagcctctggtgaagcc 843
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 853 CCCATTGTCATATAGCTTCACGACGATACACAGATATCTTCTTCCAACTTTAAGATTTCA 912
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 844 atgacccgtaagggttaacattaccctctgaagcggtacccaagggccagattcgtctgaac 903
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 913 ATGATATAAATGGGTATATTTGGAGTGTGATGATGATGATGATGATGATGATGATGAT 972
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 904 tgcagagtggtcaacagcaactc 926
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 973 TGTAAATTTTGTGAATGAGACTC 995
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

## RESULT 9

```

V59077
ID V59077 standard: cDNA: 1244 BP.
AC V59077;
DE 02-FEB-1999 (first entry)
DE Soybean seed coat peroxidase cDNA.
KW Peroxidase; seed coat; soybean; transgenic plant; ss.
OS Glycine max cv. Harosoy 63.
FH Key Location/Qualifiers
FT CDS 1..1059
FT FT /*tag= a
FT FT 1..78
FT FT /*tag= b
FT FT mat_peptide 79..1056
FT FT /*tag= c
FT FT primer_bind complement (53..68)

```

```

FT FT /*tag= d
FT FT /note= "primer prx9+"
FT FT complement (143..162)
FT FT /*tag= e
FT FT /note= "primer prx12+"
FT FT 254..273
FT FT /*tag= f
FT FT /note= "primer prx10-"
FT FT complement (280..299)
FT FT /*tag= g
FT FT /note= "primer prx2+"
FT FT 885..904
FT FT /*tag= h
FT FT /note= "primer prx6-"
PN CA2186833-A.
PD 31-MAR-1998.
PE 30-SEP-1996; 186833.
PR 30-SEP-1996; CA-186833.
PA (MIAC ) CANADA MIN AGRICULTURE.
PI G1|zen M;
PI WPI: 98-521791/45.
DR P-PSDB: w73138.
DR New DNA encoding soya seed-coat peroxidase and the promoter - useful
PT for expression of heterologous genes in the seed coat, which may
PT give herbicide resistance, viral protection and pest and pathogen
PT control, and may improve the taste or nutritional value of soya
PT beans.
PS Claim 1. Fig 1: 61pp: English.
CC This nucleotide sequence includes an open reading frame coding for
CC a seed coat peroxidase (see W73139) of soybean. To isolate the
CC peroxidase transcript, a cDNA library was constructed from
CC developing seed coat tissue of the Epep cultivar Harosoy 63, and
CC was screened with a 17-mer oligonucleotide probe (see V59079)
CC corresponding to the conserved active site domain of plant
CC peroxidases. 6 Clones contained identical cDNA transcripts. A
CC genomic clone (see V59078) was also isolated and shown to contain 4
CC exons. The 5' regulatory region of the gene is also claimed. The
CC heterologous expression, as well as vectors and host cells to be
CC used for the expression of the seed coat peroxidase, are disclosed.
CC Probes derived from the cDNA or genomic DNA can be used to detect
CC polymorphisms that distinguish Epep and eep genotypes. The
CC seed-specific regulatory region can be used to control expression
CC of genes of interest within the seed coat of transgenic plants.
SQ Sequence 1244 BP; 358 A; 261 C; 250 G; 375 T;

```

Query Match 25.9%; Score 239.8; DB 1; Length 1244;

Best Local Similarity 55.3%; Pred. No. 4.4e-68;  
Matches 510; Conservative 0; Mismatches 407; Indels 6; Gaps 2;

```

QY 4 cagttaacccctacattctcagacaatagctgtcccaagtgctccaacatcgctgcgac 63
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 79 CAGCTTACTCTCTAGTCTGACAGAGAAACATGTCCAAATGTTCCTATTGGTTTGCA 138
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64 acaatcgtaacagcgtcagatccgacatccagagatcgctgtccaatatatcgtctgcac 123
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 139 GTATATCTTCGATGCTTCTTTCACCGATCCCGAATGCGGGCCAGTCTCATAGAGCTTAT 198
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 124 ttccatgactgctctgtaaggttgcgacgtagacatatactactgagcaaacaccacagt 183
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 199 TTTCATGATGATTGCTTGTCAAGGTGTGATGATGATGATGATGATGATGATGATGATGAT 258
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 184 ttccgcacgtaaaagatgattcgggaagcgttaacagcgccagggtccttcagtgatc 243
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 259 ATAGAAAGCGAGCAAGATGATGCTTCACAAATATACATTAAGAGATTTGGAGTGTGC 318
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 244 gatcgatgaagcctgcgcttgatcgacgacgtgccacgaacagtcagttgtgcagccctg 303
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 319 AATGACATCAAGACAGCGGTGGGAAATAGTTGTCCAGACACAGTTCCTTGTGCTGATATT 378
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 304 ctgacataagctgcgcaacagagcgtgactctgacgcgagcgagcgtctcgtgagagtgccg 353
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 379 CTTCGATATTGACGTGAATATAGCTTCGTTCTGGAGAGAGCTCCAGAGATGGCCAGTTCCA 438
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```
QY 364 ctgcgtgcagctgtaactccctacaggaatccttagatctgcgcgaacgccaacttgctgct 423
    || || || || || || || || || || || || || || || || || || || || || ||
D 439 TTAGGAAGAAGGAGACAGCTTAACAGCAAAACCGTTCGAATCAAAACCTTCAGCA 498
QY 424 cactctcaccctcgcgcgcgcgcgaaggaatagctttagaagaactggtgctgcatcgtcg 483
    || || || || || || || || || || || || || || || || || || || || || ||
D 499 CTTTCTTCAACTCACTCACTTAAGGCTTCCTTGTCTTCAAGGCTCCAA--CACC 555
QY 484 agtgacctgtgctcgtcgcgagagacacatttggaagaacagtgtagtctatc 543
    || || || || || || || || || || || || || || || || || || || || || ||
D 556 CTGATTGATTACACTCTAGCTGATAGCTTTGGAAAGAGCTCGGTGATACATTTC 615
QY 544 atggataagctctacaaattcagaacaactgggttlaacttgaccccaagctgaaactag 603
    || || || || || || || || || || || || || || || || || || || || || ||
D 616 ATAAACCGATTATATCAACTCTACAGCAACACTGAAACCCCTGATCCAACTGCAACACACA 675
QY 604 taltccagaacttgagagagctgtgcgccactgaatgagcaactcagtgactagtgagac 663
    || || || || || || || || || || || || || || || || || || || || || ||
D 676 TACTTGAAGATTATGCTGCAAGATGCCCCCAAGATGCAACTGGGATTAACCTCACCACAT 735
QY 664 ttgatactgcgagaccacacatcttcgaatacaagtaactgtgaactagaggaagcag 723
    || || || || || || || || || || || || || || || || || || || || || ||
D 736 TTGGACCTGAGCACACCTGATTCATTTGACACAGATACTACTCCAACTCTTCTGCACTC 795
QY 724 aaagccctgatacagagtgatcaagaactgtttagcagtcgacagccacactgacaccatc 783
    || || || || || || || || || || || || || || || || || || || || || ||
D 796 AATGCGCTTACTTCAAGAGTACCAAGAACTTTCTCCACTCTCGTG--CTGATACCATTT 852
QY 784 cccctgtgtggaagcttctgtaactctactcaaaccttctttaaagcctctgtggaagcc 843
    || || || || || || || || || || || || || || || || || || || || || ||
D 853 CCAATGTCATATAGCTTCACAGCTTAACCAAGATATCTTTTCCAACTTGAAGTTTCA 912
QY 844 atgacccgatagtgataacttaccctctgaacggtgaacgaagccagactgctgtaac 903
    || || || || || || || || || || || || || || || || || || || || || ||
D 913 ATGATATAAATGSGTATATTTGAGATGTCGACCTGGGATGAAGAGAAATTCGCTTGCAA 972
QY 904 tgcagaagtgatcaacagcaactc 926
    || || || || || || || || || || || || || || || || || || || || || ||
D 973 TGTAAATTTGTGTAATGAGACTC 995

RESULT 10
V63720
ID V63720 standard; cDNA; 1355 BP.
AC V63720;
DT 12-APR-1999 (first entry)
DE Maize per5 root preferential cationic peroxidase cDNA.
KW Peroxidase; per5 gene; maize; corn; transgenic plant; promoter;
KW root; ss.
OS Zea mays.
FH Key Location/Qualifiers
FT CDS 53..1054
FT FT /*tag= a
FT FT sig_peptide 53..148
FT FT mat_peptide 149..1051
FT FT /*tag= b
FT FT /*tag= c
FT FT polyA_signal 1293..1298
FT FT /*tag= d
PN W09856921-A1.
PD 17-DEC-1998.
PE 10-JUN-1998; 011921.
PR 12-JUN-1997; US-049752.
PA (DOWC) DOW AGRSCIENCES LLC.
PI Ahnley M, Armstrong K, Belmar S, Folkerts O, Hopkins N,
PI Menke MA, Paredy D, Petolino JF, Smith K, Woosley A;
DR P-PSDB: W87893.
DR New isolated regulatory sequences for transgenic plants - which are
PT derived from the maize root preferential cationic peroxidase protein
PS (per5) gene
Example 3; Page 91; 150pp; English.
```

CC This is the nucleotide sequence of a maize per5 cDNA clone encoding  
CC root preferential cationic peroxidase (see W87893). The clone was  
CC isolated by screening maize root, stem, leaf, kernel and tassels RNA  
CC using probes (see V63718-19) based on tobacco peroxidase conserved  
CC domains. A probe generated from per5 cDNA was used to isolate the  
CC per5 gene (see V63717). Regulatory sequences of the per5 gene,  
CC including the promoter, introns and 3' untranslated region (3'UTR),  
CC are used in claimed recombinant gene cassettes for controlling  
CC expression of recombinant genes in selected tissue, especially the  
CC root, of transformed plants, particularly maize.

Sequence 1355 BP; 333 A; 344 C; 321 G; 351 T; 6 U;

Query Match 24.9%; Score 230.4; DB 1; Length 1355;

Best Local Similarity 54.8%; Pred. No. 5; 2e-65;

Matches 503; Conservative 0; Mismatches 406; Indels 9; Gaps 2;

```
QY 7 ttaacccttaacttcaagacaatagctgcccaagctgtccaactcgctgcgacaca 66
    || || || || || || || || || || || || || || || || || || || || || ||
D 140 TTGTTCCCAAGTTCATATGCAATGCTGCTCCCAAGGCAAGAGATTGTGCACTCCATT 199
QY 67 atcgtaacagctcagatccgataccagatcgctgctcaatatatagctgcacctc 126
    || || || || || || || || || || || || || || || || || || || || || ||
D 200 GTGGCAAGGCTGTGGCCAAAGAGACAGAGATGGCGCATTTTACTGCACTGCAATTC 259
QY 127 catgactgtcttgtaatggtltgcagcgttagcatactaaactgacacacacacagttc 186
    || || || || || || || || || || || || || || || || || || || || || ||
D 260 CANGACTGCTTTTCAAGGGCTCGCATGCTGCTGCTGTTGACAAACAGACACACATTA 319
QY 187 cgcactgaaagaatgacatcctgggaagcgtlaacagcgcgaagggcttccagtgatcat 246
    || || || || || || || || || || || || || || || || || || || || || ||
D 320 GTTAGTAGAAGAGGTCACCAACCCAGAACAGAACTCCCTCAAGGGGCTTGGAGTAGCAC 379
QY 247 cgcataagagctgcgtgtgtagtgcatacccaagacagtcagtgatgacagccgcgcg 306
    || || || || || || || || || || || || || || || || || || || || || ||
D 380 CAAATTAAGGCTGCTCTTGAAGCTGCTGCTGCCAGGACAACTCTCTGTGCGGCAATTTGT 439
QY 307 actatagctgcgcaacagagcgtgtaactctgcagagcggaacgctcgtggaagtgacgc 366
    || || || || || || || || || || || || || || || || || || || || || ||
D 440 GCCCTTGGGGCTGCTGATTTCCACCGCCCTGGTGGTGGAGACAAATAGTGGAGCGTCACTT 499
QY 367 gctcgaagtgactccctacaagcattcctagatctgcaacgcgaactgctgctcca 426
    || || || || || || || || || || || || || || || || || || || || || ||
D 500 GGCCGAGAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 559
QY 427 tcttaccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 486
    || || || || || || || || || || || || || || || || || || || || || ||
D 560 AACAAACACACTCCCTCACTATCATACCAAGTTCAAGCGCCAGGCGCTCAATGTG--TT 616
QY 487 gacctgtgtctctgtcgcgagagacacacattggaagaacacagtgtagtcatcatc 546
    || || || || || || || || || || || || || || || || || || || || || ||
D 617 GATGTTGTGCGCCCTCTCAGGTGTCACACCATGATGATGTCGTCGTCGTCGTCGTCGTCG 676
QY 547 gataagcttacaatttcagaacacactggttactgaccccaagcgtgaacactagat 606
    || || || || || || || || || || || || || || || || || || || || || ||
D 677 CAGAGGCTATACAAACAGACAGGCAATGCAATGCTGCAAGCAGCACTGATGATATCTAC 736
QY 607 ctccagacactgagagcgtgtgtgcccactgaatggaactgaatgacatagtgagactt 666
    || || || || || || || || || || || || || || || || || || || || || ||
D 737 GCCGCAAGCTGAGGAGGAGATGCCCCCGCTGTGGTGAGCAACAACCTTCCCTCTTG 796
QY 667 gatctgcgagaccacaacacatctcgaatacaagtaactatgtaactagagagacaaa 726
    || || || || || || || || || || || || || || || || || || || || || ||
D 797 GACTTCATCACCCCTGCGCAAGTTTGACAAATTTTACTACAGAACTCTCTGGCGGCAAG 856
QY 727 gacctgatacagagtgatcaagaactgtttagcagctcagacgcgaactgacacatcca 786
    || || || || || || || || || || || || || || || || || || || || || ||
D 857 GGCCTTCAAGCTCTGATAGATTCTGTAAACCA-----AGAGCGCTGACAGCGGCC 910
QY 787 ctggtgagaagtttctgaactctactcaaaccttctttaaagccttgtagaagcag 846
    || || || || || || || || || || || || || || || || || || || || || ||
D 911 CTCGTGAAGGCAATGCTGCTGATGTCAATCTCTCTCCAGACACTTTCACAGACTATG 970
```

QY 847 gaccgatggaatacattaccctctgacgggatacccaaggccagatctctgaactgc 906  
+ ||||| ||||| || ||||| || ||||| ||||| ||||| ||||| |||||  
Db 971 GTGAATATGGGAACATCTCCGCACTGACAGGCTCACAAGGTAGATCAGAGAAGAACTGCC 1030  
QY 907 agatggtcaacagcaac 924  
+ || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1031 AGGAGGCTCAACATGTAC 1048

## RESULT 11

087101  
ID 087101 standard; DNA; 1270 BP.  
AC 087101;  
DT 01-DEC-1995 (first entry)  
DE The Tpa gene.  
KW Sea Island Cotton; fibre-specific promoter; strength; peroxidase;  
KM transgenic; ds.  
OS Synthetic.  
PN MO9508914-A.  
PD 06-APR-1995.  
PE 28-SEP-1994; U11121.  
PR 30-SEP-1993; US-130086.  
PA (CETU ) AGRACETUS INC.  
PI Maliyaka J.  
PI WPI: 95-147174/19.  
PT Transgenic cotton plant expressing heterologous peroxidase -  
PT having increased fibre strength, and improved characteristics for  
PT use in the textile industry  
PS Disclosure; Page 63-64; 78pp; English.  
CC A peroxidase gene expression vector may be constructed which  
CC contains one or several peroxidase genes and a fibre-specific  
CC promoter (087082 and 087083). To construct this vector the coding  
CC region(s) of tobacco peroxidase (Tpa) gene(s) are fused to fibre-  
CC specific promoters in such a way that transcription will occur from  
CC an untranslated leader sequence in the 3' end of the promoter (087084).  
CC The expression vector can be incorporated into a phage and used to  
CC produce a proposed transgenic cotton plant capable of producing fibres  
CC of increased strength.  
CC The invention is useful in producing strong cotton fibres leading to  
CC stronger fabrics. Increased fibre strength allows textile manufacturing  
CC processes, e.g. spinning and weaving, to be carried out at higher speeds  
CC resulting in lower manufacturing costs. The technique used may also be  
CC to change other of the transgenic plants characteristics e.g water  
CC absorption.  
SQ Sequence 1270 BP; 390 A; 195 C; 246 G; 439 T;

## Query Match

Best Local Similarity 17.6%; Score 163.4; DB 1; Length 1270;  
Matches 481; Conservative 0; Mismatches 421; Indels 15; Gaps 5;

QY 4 cagttaccctcatctcagacaatagctgtcccaagctgtccaacatgctcgagac 63  
+ || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 81 CAATTAAAGCGCAATTTTATGATACACCTTCCCTTAATGTACAGATTTGACGTGT 140  
QY 64 acaatgctcaacggctcagatccatcccaaggatcgctgtcaatataatcagctgcac 123  
+ || || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 141 GTTATGATCAAGGCAACGACTGATGCTCGACCTGGTCTAAATTAATTCGCTTCAT 200  
QY 124 ttccatgactgactcgtgaatgctgtcgacgctagacatattactggaacaacacagct 183  
+ || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 201 TTCCATATGATTGTTTGTAAATGTTGATGATGATCAATTTTGTAG--ACACAGATGGG 257  
QY 184 ttccgcactgaagaagatgcatctgggaacgctaacagcgccagggccttcagatc 243  
+ || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 258 ACTCAAACTGAGAAAGATGCACTGCTAA--TGTAGGTGAGAGGATTTGATATTGTG 314  
QY 244 gatcgcatgaaagctgctgtgagctcaagcatgcccagaagaagcagctgtgacagctg 303  
+ || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 315 GATATATTTAAACTGCACTAGAAATGTATGCCCTGGTGTGATCTTGCGAGATATT 374  
QY 304 ctgactatagctgcgcaacagagagctgactcttcgcaaggcgagacgctctggaagtgccg 363  
+ || || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 375 TTACCCCTTGACATCTGAATTTGGAGTTGTCTTGCTTAAGGCTCCGTGGCAAGTACTT 434  
QY 364 ctgagtgagtgatcccttaacagagcatctctagatctgcccacagcactgtccgct 423  
+ || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 435 TTTCGCAAGAGATPACCTTAACGCAAAACGATCTGAGCTAATGATATTCCTTAGC 494  
QY 424 ccattcttcaccctgcgcccaagctgaagatagctcttgaacagctgggtcgaactgc 483  
+ || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 495 CCCTTTGAACCCCTTGCTGTATATGACCCCAATTCACCAAGGCAATTCGAT--TTA 551  
QY 484 agtgacctgtgctctgtccggaagacaacatcttgaaagaaacagtgatgctc 543  
+ || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 552 ACTGATCTGTTGCTTATCTATGTGTGACACACATTTGGAAGAGCAAGATGTGACTTTT 611  
QY 544 atgatatggtctcaatttagcaaacactgtgtactctgcccccaagctgaaactcag 603  
+ || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 612 GAACACGCTCTTTTAACCTTACGTGAGAGTGATCTGATCCTGATCCAGCCGAGACCTACA 671  
QY 604 tatcccaagacactgaagagctgtgcccactgaatg---caaccctcaagtacagtg 660  
+ || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 672 TTTTTCACAAACATTAACAGGCAATTTGCTTCACAGTGGAAATATGGCAATCTTTTACA 731  
QY 661 gacttgatctgcggaaccccaacatctctgataacaagtaactatgtgaactagagag 720  
+ || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 732 AATCTGTATATATCACTCTAATGATTTTGAATGATGACATTTTCACTAATCTCAAAAT 791  
QY 721 cagaaagccctgatacagatgtatcaagaactgtttgacagtcgaagccactgaacc 780  
+ || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 792 AATCAAGACCTCTTTCAAACGATCAAGAGTTGTT--TCTACATCTGATCTGCACACA 848  
QY 781 atcccactggtgaagaaattttgtcaactctactcaaaccttcttaacgaccttgagaa 840  
+ || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 849 ATTCTATATAGTAATCGTATATGCTGTAGTCAAGTCAGTTTGTGATGATTTATTTTGC 908  
QY 841 gccatgaccglatggtlaacattaccctctcgacgggttacccaagggccagatctgtg 900  
+ || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 909 TCGATGATTAAATGGGTAATATATAGTCCATTACTGTACTAATGAGAGATTAGGAAA 968  
QY 901 aactgcagatgtgtcaa 917  
+ || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 969 GATTGCAAGAGGTTAA 985  
+ || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

## RESULT 12

T63685  
ID T63685 standard; cDNA; 1270 BP.  
AC T63685;  
DT 11-JUN-1997 (first entry)  
DE Full-length Tpa-1 gene.  
KW primer; PCR; polymerase chain reaction; cotton fibre; seed floss fibre;  
KW promoter; peroxidase; production; fibre strength; Sea Island;  
KW Gossypium sp.; Coker; Kapok; G. barbadense; ds.  
OS Nicotiana tabacum.  
PN US5608148-A.  
PD 04-MAR-1997.  
PE 30-SEP-1993; 130086.  
PR 30-SEP-1993; US-130086.  
PR 25-JAN-1995; US-378588.  
PA (CETU ) AGRACETUS INC.  
PI John ME;  
PI WPI: 97-164559/15.  
PT Transgenic cotton plants with increased fibre strength - are  
PT transformed with construct containing peroxidase gene  
PS Example D; Column 45-46; 43pp; English.  
CC The present sequence is the full-length Tpa-1 gene which expresses the  
CC tobacco anionic peroxidase. Cotton plants whose genome contains  
CC heterologous genetic construct comprising a seed floss fibre (SFF)-  
CC specific promoter isolated from cotton plants and a coding sequence  
CC encoding a peroxidase are useful for production of cotton with increased  
CC fibre strength. The peroxidase genes are over-expressed in fibre and not  
CC in other plant tissues where it would be harmful to the plant.  
SQ Sequence 1270 BP; 390 A; 195 C; 246 G; 439 T;



[illegible]

AC	X03970:1	(first entry)
DT	09-APR-1999	Full Length anionic tobacco peroxidase gene Tpa-1.
KM	Cotton; seed; fibre; promoter; cotton peroxidase; floss; fabric; ds.	
OS	Nicotiana tabacum.	
PN	US5869720-A.	
PD	09-FEB-1999.	
PF	03-MAR-1997; 811094.	
PR	25-JAN-1995; US-378588.	
PR	30-SEP-1993; US-130086.	
PR	03-MAR-1997; US-811094.	
PA	(MONS ) MONSANTO CO.	
PI	John ME; WPI; 99-152903/13.	
PT	Transgenic cotton seed floss fibre with increased strength -	
PS	containing peroxidase gene under control of fibre-specific promoter	
CC	Example; Column 45-46; 44pp; English.	
CC	The present invention describes a cotton seed floss fibre whose genome	
CC	contains a heterologous construct comprising a peroxidase coding	
CC	sequence linked to a seed floss fibre-specific promoter. The present	
CC	sequence represents the anionic tobacco peroxidase gene Tpa-1, from an	
CC	example of the present invention. The fibre has greater strength than	
CC	a corresponding untransformed fibre, so such fibres can be made into	
CC	fabrics at higher machine speeds and produce harder-wearing fabrics.	
SC	Sequence 1270 BP; 390 A; 195 C; 246 G; 439 T;	
Query Match	17.6%; Score 163.4; DB 1; Length 1270;	
Best Local Similarity	52.5%; Pred. No. 3e-43;	
Matches	481; Conservative 0; Mismatches 421; Indels 15; Gaps	
OY	4	cagttaaccctcatattctacgacataagctgltcccaacgltgccaaacgltcgcgac 63
DB	81	CAATTAAAGGGCAACATTATACGATACCACTTGCCCTAATGTTAACAAGTATGACGTGT 140
OY	64	acaatcgcacaagagctcagatccgataccagagatcgctgtcctaataatlaagctgcac 123
DB	141	GTTATGAGATCAAAAGCCAACTGATGCTCAGCGTGGGTCTAAATTAATTCCTTCAT 200
OY	124	ttccatagctgcttcgtgaatgttttgagagctgatactatcttgagcaacaacacact 183
DB	201	TTCCATGATTTGTTTGTTATATGCTGTGTGATGATCAATTTTGTAG--ACACAGATGG 257
OY	184	ttccgactgtaaaagatgatacttcgggaagcgttaacagccagcagggtcttccagtgac 243
DB	258	ACTCAACTGAGAAAGATGACCTGCTA--TGATGGTGCAGGAGATTTGATTTGTG 314
OY	244	gatcgcatgaagagctgcggttgatgcagatgcccagcaagtcagtgtgtcagacctg 303
DB	315	GATGATATTAAACGCACTAGAGATGATCCCTGGTGTGTCTTGTGTCCACATATT 374
OY	304	cgaatatagctcgcgcaacagagcgltgactcttcgaagggagacgctcccgagagtgccg 363
DB	375	TTTAGCCCTTGCACTCGAANAATTTGGAAGTTGCTTTGGCTAAAGGTCCTGTCGCAAGTACTT 434
OY	364	ctcggctcagctgatactcctacaggaacttctagatcttgccaaagcccaactgtcgtc 423
DB	435	TTTGGCAGAGAAGATGAGTTAAGACCAACGATCTGGAGCTAATATGATATTCTCTAC 494
OY	424	ccattcttaaccctgcgcccaagctgaagataagactttagaagacgtggtgtgaatcgctcg 483
DB	495	CCCTTTGGAACCCCTTGCTGTAATGACGCCACAAATTCACCAACAGAGATGAT--TTA 551
OY	484	agtgaccttgggctcgtgcggaagaacaacatttgtaaaaaaacacagtgatgtatc 543
DB	552	ACTGATCTTGTTGGCTCTATCTGCTGCACACAAATTTGGAAGCAAGATGGTACTTTT 611
OY	544	atgataagctctacaatttcagcaaacactgggttaccctgaccccaagctgtaaacactag 603
DB	612	GAACAAGCTCTTTTAATTCTCAGTGGCAAGTGAATCTCTGATCCAAACCTTAAGCCTTACA 671
OY	604	tatcccaagacactgagaagctgtgcccagaatg---caactcagtgactagtg 660





PF 27-JUL-1992; G01389.  
 PR 29-JUL-1991; GB-016325.  
 PR 05-FEB-1992; GB-002401.  
 PA (BRBI-) BRITISH BIO-TECHNOLOGY LTD.  
 PI Bawden L, Edwards RM;  
 DR WPI: 93-076509/09.  
 DR P-PSDB; R32761.  
 PT New Insulin-like Growth Factor II analogues - for treating growth  
 disorders, osteoporosis and muscle wasting diseases, have less  
 PT affinity for insulin receptors than unmodified IGF-II  
 PS Example 4: Page 50; 71pp; English.  
 CC This sequence encodes a fusion protein comprising the first 53  
 CC amino acids of HRP fused to IGF-II via a unique methionine residue  
 CC to allow for the release of mature IGF-II by cleavage with CNBr.  
 CC The fusion sequence is better for expression of IGF-II than a  
 CC fusion of the first 282 amino acids of HRP with IGF-II (see 037614)  
 CC in which IGF-II is only 19% of the total fusion protein and the HRP  
 CC sequence contains two internal methionine residues which act as  
 CC additional cleavage sites. The IGF-II is the wild-type sequence,  
 CC i.e. with Arg residues at positions 37 and 38. The invention covers  
 CC analogues of this sequence in which one or both of the Arg residues  
 CC are substid. by other amino acids (pref. non-basic and esp. Gln).  
 CC The mutants are protease-resistant and have reduced affinity for  
 CC insulin receptors.  
 SQ Sequence 388 BP; 85 A; 113 C; 90 G; 100 T;

Query Match 17.2%; Score 159.4; DB 1; Length 388;  
 Best Local Similarity 99.4%; Pred. No. 3,3e-42;  
 Matches 160; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgcagltaacccctacatctctacagacaatagctgtcccaagctgtccaaatcgtcgc 60  
 |||||||  
 DB 11 ATGCAGTTAAACCCCTACATTCTACGACAAATAGCTGTCCCAACGTGTCCAAATCTTCGC 70  
 |||||||  
 QY 61 gaccacaatgcgaacggagcccaatccagatccgctgcttcaatattagctcgtc 120  
 |||||||  
 DB 71 GACACAAATCGTCAAGAGCTCAGATCCGATCCAGATCGCTGCTCAATATTACGTCTG 130  
 |||||||  
 QY 121 cacttccatgactgctcgtgaatggtgcgaagctagcat 161  
 |||||||  
 DB 131 CACTTCATGACTGCTCTCGTAATGTTGCGACCGCTAGTAT 171  
 |||||||

Search completed: October 4, 2000, 06:40:45  
 Job time: 12254 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 05:49:03 ; Search time 74.16 Seconds  
(without alignments)  
1890.518 Million cell updates/sec

Title: US-09-246-451-16

Perfect score: 927  
Sequence: 1 atgcagtaacctacatc.....gagtgtcacagcaactct 927

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgnl\_7/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgnl\_7/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgnl\_7/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgnl\_7/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgnl\_7/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgnl\_7/ptodata/1/ina/PCTUS\_COMB.seq:\*  
7: /cgnl\_7/ptodata/1/ina/backfilist1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	843.8	91.0	1072	2	US-08-190-029A-9
2	843.8	91.0	1072	3	US-08-462-695-9
3	163.4	17.6	1270	1	US-08-378-588-23
4	163.4	17.6	1270	3	US-08-811-094-23
5	163.4	17.6	1270	6	PCT-US94-11121-23
6	159.4	17.2	388	2	US-08-190-029A-11
7	159.4	17.2	388	3	US-08-462-695-11
8	148.6	16.0	1131	1	US-08-181-271A-22
9	148.6	16.0	1131	1	US-08-449-315-22
10	148.6	16.0	1131	1	US-08-444-803-22
11	148.6	16.0	1131	1	US-08-449-043-22
12	148.6	16.0	1131	2	US-08-456-265A-22
13	148.6	16.0	1131	2	US-08-455-416-22
14	148.6	16.0	1131	2	US-08-455-244-22
15	148.6	16.0	1131	2	US-08-454-876-22
16	148.6	16.0	1131	2	US-08-457-364-22
17	148.6	16.0	1131	3	US-08-456-262-22
18	148.6	16.0	1131	3	US-08-456-240-22
19	148.6	16.0	1131	3	US-08-455-736-22
20	148.6	16.0	1131	4	US-08-971-217-22
21	146	15.7	1181	1	US-08-181-271A-27
22	146	15.7	1181	1	US-08-449-315-27
23	146	15.7	1181	1	US-08-444-803-27
24	146	15.7	1181	1	US-08-449-043-27
25	146	15.7	1181	2	US-08-456-265A-27
26	146	15.7	1181	2	US-08-455-416-27

27	146	15.7	1181	2	US-08-455-244-27	Sequence 27, Appl
28	146	15.7	1181	2	US-08-454-876-27	Sequence 27, Appl
29	146	15.7	1181	3	US-08-457-364-27	Sequence 27, Appl
30	146	15.7	1181	3	US-08-456-262-27	Sequence 27, Appl
31	146	15.7	1181	3	US-08-456-240-27	Sequence 27, Appl
32	146	15.7	1181	3	US-08-455-736-27	Sequence 27, Appl
33	146	15.7	1181	4	US-08-971-217-27	Sequence 27, Appl
34	129	13.9	1064	1	US-08-378-588-15	Sequence 15, Appl
35	129	13.9	1064	3	US-08-811-094-15	Sequence 15, Appl
36	129	13.9	1064	6	PCT-US94-11121-15	Sequence 15, Appl
37	119.8	12.9	1326	3	US-08-671-320-12	Sequence 12, Appl
38	119.8	12.9	1326	3	US-08-868-577-12	Sequence 12, Appl
39	107	11.5	1167	3	US-08-671-320-16	Sequence 16, Appl
40	107	11.5	1167	3	US-08-868-577-16	Sequence 16, Appl
41	105.4	11.4	1314	3	US-08-868-577-10	Sequence 10, Appl
42	105.4	11.4	1315	3	US-08-671-320-10	Sequence 10, Appl
43	98.2	10.6	1191	3	US-08-671-320-14	Sequence 14, Appl
44	98.2	10.6	1191	3	US-08-868-577-14	Sequence 14, Appl
45	79.8	8.6	686	1	US-08-181-271A-28	Sequence 28, Appl

#### ALIGNMENTS

RESULT 1  
US-08-190-029A-9  
; Sequence 9, Application US/08190029A  
; Patent No. 5736363  
; GENERAL INFORMATION:  
; APPLICANT: EDWARDS, Richard Mark  
; APPLICANT: BARDEN, Lindsey  
; TITLE OF INVENTION: IGF-II ANALOGUES  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: ALBERETTI & WITCOFF, LTD.  
; STREET: 10 S. WACKER DRIVE, SUITE 3000  
; CITY: CHICAGO  
; STATE: ILLINOIS  
; COUNTRY: U.S.A.  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/190,029A  
; FILING DATE: 28-FEB-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB92/01389  
; FILING DATE: 27-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9202401.7  
; FILING DATE: 05-FEB-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9116325.3  
; FILING DATE: 29-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: JOHN J. MCDONNELL  
; REGISTRATION NUMBER: 26,949  
; REFERENCE/DOCKET NUMBER: 94,062  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1072 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:

```

? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 11..1057
? OTHER INFORMATION: /codon_start= 11
? OTHER INFORMATION: /function= "Gene for IGF-II/HRP fusion protein"
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 1..6
? OTHER INFORMATION: /function= "HindIII cleavage site"
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 1067..1072
? OTHER INFORMATION: /function= "EcoRI cleavage site"
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 11..856
? OTHER INFORMATION: /function= "HRP coding sequence"
? OTHER INFORMATION: (fragment)"
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 857..1057
? OTHER INFORMATION: /function= "IGF-II coding sequence"
? US-08-190-029A-9
```

Query Match 91.0%; Score 843.8; DB 2; Length 1072;

Best Local Similarity 99.8%; Pred. No. 4.9e-277;

Matches 845; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 atgcagctaacccctacatctctacagcaatagctgtcccaagctgtcccaacatcgttcgc 60
DB 11 ATGCAGCTTAACCCCTCATCTTCTACGACAAATAGCTGTCCCAAGCTGTCCCAATCGTTCCG 70
QY 61 gacacaacgtlcaagagctcagatccagatcccaagatcgtcttcaatattacgtctg 120
DB 71 GACACAACGTCTCAACGAGCTCAGATCCGATCCAGATCCGCTCTCAATATTACGCTG 130
QY 121 cacttccatgacgtctgtggaatgtgtgagagctagcattacttgcgaacaccacc 180
DB 131 CACTTCCATGACTGCTTGTGGAAATGCTGTGAGCTGATATTACTGGACAACACCACC 190
QY 181 agttccgacactgaaagagatgcatcgcggaagcgaacgcgcgaagggcgttccagtg 240
DB 191 AGTTCCGACCTGAAAGAGATGCAATTCGCGAAGCTAAGCGCCGAGGGGCTTCCAGTG 250
QY 241 atcgalcgaatgaagagctgcgttgagtcagcaltgcccaagaaacagltgtlgaagac 300
DB 251 ATCGATCGCATGAAGGCTGCGCTTGAGTCAGCATGCCACGAACAGTCAAGTTGTGCAGAC 310
QY 301 ctgtctactatagctgcgcaacagagcgtgactcttcgagggcgagaccgttctgagagtg 360
DB 311 CTGTCTACTATAGCTGCGCAACAGACGCTGACTCTTGCAAGCGGACCGTCTGGAAGTG 370
QY 361 ccgctcgltcgaagcttactccctacagagcaltcctgaltctggccaagccaacttgct 420
DB 371 CCGCTCGGTCGAGCTGACTCTTACAGGCAATTCCTAGATCTGGCCAAAGCGCAACTTGCT 430
QY 421 gctcacttctcaaccctgcgcagctgaagatagctttagaagacgtlgaagtcgc 480
DB 431 GCTCATTCTTCACCTGCGCCAGCTGAAGGATAGCTTTGAAACGTGGGTTGAATCGC 490
QY 481 tggagagactctgtgtctgtctgcggaagacacactttggaaagaaacagtgtagttc 540
DB 491 TGGAGAGACTCTGTGTGCTGTCTGGAGGACACACTTTGGAAAGAACCAAGTGTAGGTTTC 550
QY 541 atcatgataagagctcacaattcagacaacacggtgttaccctgacccaagcgtgaagact 600
DB 551 ATCATGATAGAGCTCTACAAATTTACAGCAACACTGGGTTACTCTACCCCAAGCTGAACACT 610
QY 601 acgtatcccgagacactgagagctgtgtcccaactgaatggaactcaagtcaagttag 660
DB 611 ACGTATCTCCAGACACTGAGAGGCTGTGTGCCCACTGAAATGGCAACCTCACTGCACTAGTG 670
QY 661 gacttgaatctgagagcccaaccatcttcgataaacaagtaagtgtgaatcagaaggag 720
DB 671 GACTTGAATCTGGAGACCCCAACCATCTTTCGATACAGTACTATGTGAATCTAGAGAG 730
QY 721 cagaagagcctgataagagtgatcaagaactgtttagcagtcacagcagcactgacacc 780
DB 731 CAGAAGGCTGTATACAGAGTGTATCAAGAACTGTTAGAGAGTCCAAACGCCACTGACACC 790
QY 781 atcccaatgttgagaagtttctgtaacttaactcaactcttcaagccttctgtgaa 840
DB 791 ATCCCAATGTGTGAAGTTTGTCTACTACTCAAACTTCTTAAAGCCTTGTGGAA 850
QY 841 gcatgag 847
DB 851 GCTATGG 857
```

## RESULT 2

US-08-462-695-9

Sequence 9, Application US/08462695

Patent No. 5854025

GENERAL INFORMATION:

APPLICANT: EDWARDS, Richard Mark

APPLICANT: BARDEN, Lindsey

TITLE OF INVENTION: IGF-II ANALOGUES

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: BANNER & ALLEGRETTI, LTD.

STREET: 10 S. WACKER DRIVE, SUITE 3000

CITY: CHICAGO

STATE: ILLINOIS

COUNTRY: U.S.A.

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/462,695

FILING DATE: 5-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/190,029

FILING DATE: 28-FEB-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB92/01389

FILING DATE: 27-JUL-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9202401.7

FILING DATE: 05-FEB-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9116325.3

FILING DATE: 29-JUL-1991

AFTORNEY/AGENT INFORMATION:

NAME: JOHN J. McDONNELL

REGISTRATION NUMBER: 26,949

REFERENCE/DOCKET NUMBER: 94,062-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-715-1000

TELEFAX: 312-715-1234

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 1072 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORGANISM: Homo sapiens

FEATURE:

```

? NAME/KEY: CDS
? LOCATION: 11..1057
? OTHER INFORMATION: /codon_start= 11
? OTHER INFORMATION: /function= "Gene for IGF-II/HRP fusion protein"
? OTHER INFORMATION: /product= "Synthetic DNA"
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 1..6
? OTHER INFORMATION: /function= "HindIII cleavage site"
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 1067..1072
? OTHER INFORMATION: /function= "EcoRI cleavage site"
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 11..856
? OTHER INFORMATION: /function= "HRP coding sequence"
? OTHER INFORMATION: (fragment)"
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 857..1057
? OTHER INFORMATION: /function= "IGF-II coding sequence"
US-08-462-695-9
```

```

Query Match          91.0%; Score 843.8; DB 3: Length 1072;
Best Local Similarity 99.8%; Pred. No. 4.9e-277;
Matches 845; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```

QY 1 atcgagtltaacccctacatctcagacaatagctgtcccaacgltgcacaacatcgttcgc 60
    |||||||
DB 11 atgcagtttaacccctacatctcagacaatagctgtcccaacgltgcacaacatcgttcgc 70
    |||||||
QY 61 gaacacatgctcaagagcctcagatccgacatcccaagagctcgtcttaataatacgtctg 120
    |||||||
DB 71 GACACATCGTCACACGAGCTCAGATCCGATCCAGATCCGCTGTTAAATTAACGCTCG 130
    |||||||
QY 121 cacttccatctacgtctcgtgaatgtgttcgacgctagcatatctacgtggaacaacacc 180
    |||||||
DB 131 CACTTCATGACACGCTCTCGTAATGCTGGAGCAGCCCTAGCATATTACGGAACACACACC 190
    |||||||
QY 181 agtttcgcacgcgaagagatgcattcgggaacgctaacagcgccagggcgttccagtg 240
    |||||||
DB 191 AGTTTCGCGACTGAAAGAGATGATTCGGAGACGCTTAACGCGCCAGGCGCTTCCAGTG 250
    |||||||
QY 241 atcgatcgatgaagagctgcgttgatgcagctgcccagaacagcttggtgcagac 300
    |||||||
DB 251 ATCGATCGCATGAAGGCTGCGGTGAGTCAAGCATGCCACGAACAGTCAGTTGCGAGAC 310
    |||||||
QY 301 ctgctgactatactgcgcaacagagcgtagctcttcagcgcgagccgctcctggaagtg 360
    |||||||
DB 311 CTGCTGACTATAGCTGGGCAACAGAGCGTAGCTTTGACAGCGGACCGCTCTGTGAGAGTG 370
    |||||||
QY 361 cgcgctggttcgaagtgatccctcctaacagcatccctagatctgcccacgcgaacttgct 420
    |||||||
DB 371 CCGCTCGGTGAGAGTGAATCCCTTAACAGCATTCCTAGATCTGGCCCAACGCCAATTTGCT 430
    |||||||
QY 421 gctccatcttcgaccccgcccgagctggaagatagctttagaagacggtggttgaatcgc 480
    |||||||
DB 431 GCTCCATTTTCACCCGCCCCAGCTGAAGATAGCTTTGAAMAAGTGGGTCTGGAATCGC 490
    |||||||
QY 481 tcgagtlactctgtgctcgttcgcgagagacacacatcttgaaagaacagtgtagtttc 540
    |||||||
DB 491 TCGAGTACCTTGCTGCTCTGCTCGGAGGACACACATTTTGAAGAAGACCACTAGAGTTTC 550
    |||||||
QY 541 atcatalgataggtctlaacatlttcagcaacacttggttacttgaccccaacgcttgaact 600
    |||||||
DB 551 ATCATGATAGGCTCTACAAATTTTCAGCAACACTGGTGTACTGACCCCAAGCTGAACACT 610
    |||||||
QY 601 acgtatctcagaacactggaaggtctgtgcccaactgaaatggcaacctcaagtgcactagt 660
    |||||||
DB 611 ACGTATCTCAGACACTGAGAGGCTGTGCGCCACTGTAAGTGAACCACTCAGTGCCTAGTG 670
    |||||||
```

```

QY 661 gaattgatctcgcgagcccaaccatcttcgataaagaactatgtgaatctagagag 720
    |||||||
DB 671 GACTTGTATCTGCGGACCCCAACCATCTTGATGATTAAGATGATGATGATGATGATGATG 730
    |||||||
QY 721 cagaagggcttgatagagatgtatcaagaactgtttagcagltccagagcccaactgacac 780
    |||||||
DB 731 CAGAAAGGCTGTATACAGAGTATGATCAAGAACTGTTAGCAGTCCMAAGCAGCAGACACC 790
    |||||||
QY 781 atcccaactgtgaagaatttgctcaactctacacaaacctctttagcgttcgtggaa 840
    |||||||
DB 791 ATCCACTGCTGAGAGAGTTTGCTAATCTATACAACTTTTAAAGCCCTTCTGTA 850
    |||||||
QY 841 gccatgg 847
    || ||||
DB 851 GCTATGG 857
    || ||||
```

```

RESULT 3
US-08-378-588-23
? Sequence 23, Application US/08378588
? Patent No. 5608148
? GENERAL INFORMATION:
? APPLICANT: John, Mallyakal E.
? TITLE OF INVENTION: TRANSGENIC COTTON PLANTS
? TITLE OF INVENTION: PRODUCING HETEROLOGOUS PEROXIDASE
? NUMBER OF SEQUENCES: 29
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Nicholas J. Seay, Quarles & Brady
? STREET: First Wisconsin Plaza, One South
? STREET: Pluckney St.,
? STREET: P.O. Box 2113
? CITY: Madison
? STATE: WI
? COUNTRY: USA
? ZIP: 53701-2113
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/378,588
? FILING DATE:
? CLASSIFICATION: 800
? ATTORNEY/AGENT INFORMATION:
? NAME: Seay, Nicholas J.
? REGISTRATION NUMBER: 27,386
? REFERENCE/DOCKET NUMBER: 11-229-9101-2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (608) 251-2484
? TELEFAX: (608) 251-9166
? INFORMATION FOR SEQ ID NO: 23:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1270 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
US-08-378-588-23
```

```

Query Match          17.6%; Score 163.4; DB 1: Length 1270;
Best Local Similarity 52.5%; Pred. No. 1.3e-45;
Matches 481; Conservative 0; Mismatches 421; Indels 15; Gaps 5;
```



```
Db 612 GAACACGCTCTTTAACTTCAGTGCAGTGTAAATCCATCCAGACGCTTACA 671
Qy 604 latccagacactgagagcgcttgcccaactgaatg---caacctcagtcactagtg 660
Db 672 TTTTACAAACATTACAGGGAAATTTGCTCCAGAGTGGAAATAATATGCACTACTTTTACA 731
Qy 661 gacttgcactgagagcccaacacatcttcgatacaagaactatgtgaactcagaagag 720
Db 732 AATCTGATATATCAACTCTTAATGATTTTGAATATGACTATTCACATACTCTCAAAAT 791
Qy 721 cagaagcgctgatacagagatgatacaagaactgttagcagtcagaagcgcaactgaacc 780
Db 792 AATCAAGACTCTTCAACTGATCAAGAGTGTGTT--TCTACATCTGATCTGCTACA 848
Qy 781 atccacatgtgagaagcttctgtaactctactcaaacctctcttaagccttcgagaa 840
Db 849 ATTGCTATATGATTCGTTATGCTGGTAGTCAAAAGTCAAGTTTGTGATTTATTTTGC 908
Qy 841 gccatgagccgtatggttaacataccctctcgaagtgatacccaaggccagatcgtctg 900
Db 909 TCGATGATTAATTTGGGTAATATATGATTCATTAAGTGTACTACTAATGAGAGATTAGGAAA 968
Qy 901 aactgcagagtggtcaaa 917
Db 969 GATTGCAAGAGGCTTAA 985

RESULT 5
PCT-US94-11121-23
; Sequence 23, Application PC/TUS9411121
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: TRANSGENIC COTTON PLANTS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: First Wisconsin Plaza, One South
; STREET: Pinckney St.,
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/11121
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 11-229-9076-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-2484
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1270 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US94-11121-23

Query Match 17.6%; Score 163.4; DB 6; Length 1270;
Best Local Similarity 52.5%; Pred. No. 1.3e-45;
Matches 481; Conservative 0; Mismatches 421; Indels 15; Gaps 5;
```

```
Qy 4 cagltaacccctacatcttcaagaacatagctgtcccaagtgltccaaactggtcgcgac 63
Db 81 CAATTAAAGCGCAACATTTTACAGATACACATTTGCCCTATATGTTACAAATATTTACGCGGT 140
Qy 64 acaatcgtaacagagatccagatccatccagaagatgcgtcctcaatatactagctgcac 123
Db 141 GTTATGGATCAAAAGCAACCTACTAGATGCTGCAGCTGTGCTAAATAATATTCGCTTCAT 200
Qy 124 ttccatgactgcttcgtaagtgtgtcgacgctagcatattactggaacaccaccagt 183
Db 201 TTCCATGATTTGTTTGTAAATGTTTGATGATGATCAATTTTGTGTTAG--ACACAGATGG 257
Qy 184 ttccgacatgaagaagatgatactcggaagcgtcaacagcgcaagcgcttccagtgatc 243
Db 258 ACTAAACTGGAAGATATGACCTGCTAA---TGTAGTGCAGAGAGATTTGATATTGTG 314
Qy 244 gatcgcatgaagcgtgcgtgtgagcatgcaagcgaacgaatcagltgtgcagacctg 303
Db 315 GATGATATTAAACCTGCACCTAGAGAAATGTAAGCCCTGCTGTATCTGTGCGAGATATT 374
Qy 304 ctgactatagctgcgcaacagagcggtgactcttgcagcgcgaccgctcctggaagtgccg 363
Db 375 TTAGCCCTTGCACTGGAATTTGGAGTTGCTTGGCTAAAGCTCGCTGCGCAAGTACTT 434
Qy 364 ctgcgtcgaagctgactccctacagcaatccctagatctcgccacgcgaacttgctgct 423
Db 435 TTGCGCAGAAAGATATGATCTTACAGCAACCGCATCTGAGCTAATATGATATTTCTTACG 494
Qy 424 ccaatcttcaaccctgcgccagctgaagatagctttagaaagctggtgtcgaatcgtcg 483
Db 495 CCCTTGAACCCCTTGGCTGTATGATGACGCCACAAATTCACCAAGAGGAAATGAT---TTA 551
Qy 484 agtgaactgtgctcgtcgtccgagagacacatcttggaaagaacagctgtagttcactc 543
Db 552 ACTGATCTTGTGCTCATCTGATGTCGACACACATTTGGAAAGCAGATGTGCTACTT 611
Qy 544 atgatatgagctctacaattcagaacacgtggtactcgaaccccaagctgaacactagc 603
Db 612 GAACACGCTCTTTAACTTCAGTGCAGTGTAAATCTGTGATCCAGCCTTAGACGCTTACA 671
Qy 604 latccagacactgagagcgcttgcccaactgaatg---caacctcagtcactagtg 660
Db 672 TTTTACAAACATTACAGGGAAATTTGCTCCAGAGTGGAAATAATATGCACTACTTTTACA 731
Qy 661 gacttgcactgagagcccaacacatcttcgatacaagaactatgtgaactcagaagag 720
Db 732 AATCTGATATATCAACTCTTAATGATTTTGAATATGACTATTCACATAATCTCAAAAT 791
Qy 721 cagaagcgctgatacagagatgatacaagaactgttagcagtcagaagcgcaactgaacc 780
Db 792 AATCAAGACTCTTCAACTGATCAAGAGTGTGTT--TCTACATCTGATCTGCTACA 848
Qy 781 atccacatgtgagaagcttctgtaactctactcaaacctctcttaagccttcgagaa 840
Db 849 ATTGCTATATGATTCGTTATGCTGGTAGTCAAAAGTCAAGTTTGTGATTTATTTTGC 908
Qy 841 gccatgagccgtatggttaacataccctctcgaagtgatacccaaggccagatcgtctg 900
Db 909 TCGATGATTAATTTGGGTAATATATGATTCATTAAGTGTACTACTAATGAGAGATTAGGAAA 968
Qy 901 aactgcagagtggtcaaa 917
Db 969 GATTGCAAGAGGCTTAA 985

RESULT 6
US-08-190-029A-11
; Sequence 11, Application US/08190029A
; Patent No. 5736363
; GENERAL INFORMATION:
; APPLICANT: EDWARDS, Richard Mark
; APPLICANT: BAWDEN, Lindsey
```

;; TITLE OF INVENTION: IGF-II ANALOGUES  
;; NUMBER OF SEQUENCES: 12  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: ALLEGRETTI & WITCOFF, LTD.  
;; STREET: 10 S. WACKER DRIVE, SUITE 3000  
;; CITY: CHICAGO  
;; STATE: ILLINOIS  
;; COUNTRY: U.S.A.  
;; ZIP: 60606  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;;  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/190,029A  
;; FILING DATE: 28-FEB-1994  
;;  
;; CLASSIFICATION: 435  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/GB92/01389  
;; FILING DATE: 27-JUL-1992  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: GB 9202401.7  
;; FILING DATE: 05-FEB-1992  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: GB 9116325.3  
;; FILING DATE: 29-JUL-1991  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: JOHN J. McDONNELL  
;; REGISTRATION NUMBER: 26,949  
;; REFERENCE/DOCKET NUMBER: 94,062  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 312-715-1000  
;; TELEFAX: 312-715-1234  
;;  
;; INFORMATION FOR SEQ. ID NO: 11:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 388 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; ORGANISM: Homo sapiens  
;;  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 11..373  
;; OTHER INFORMATION: /codon\_start= 11  
;;  
;; OTHER INFORMATION: /function="gene for IGF-II/HRP fusion protein"  
;;  
;; OTHER INFORMATION: /product="Synthetic DNA"  
;;  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: 1..6  
;; OTHER INFORMATION: /function="HindIII cleavage site"  
;;  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: 383..388  
;; OTHER INFORMATION: /function="EcoRI cleavage site"  
;;  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: 11..172  
;; OTHER INFORMATION: /function="HRP coding sequence"  
;;  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: 173..373  
;; OTHER INFORMATION: /function="IGF-II coding sequence"  
;;  
;; US-08-190-029A-11

Query Match 17.2%; Score 159.4; DB 2; Length 388;  
Best Local Similarity 99.4%; Pred. No. 1.4e-44;  
Matches 160; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgcaagttaacccctacatctctacgacacatagctgtcccaacgltgcacacatcgttcgc 60

Db 11 atgcagttaacccctacatctctacgacacatagctgtcccaacgltgcacacatcgttcgc 70  
QY 61 gacacatcgtcacagcagcctcacatccgatccacagatcgtcttcaataatcgtc 120  
Db 71 gacacatcgtcacagcagcctcacatccgatccacagatcgtcttcaataatcgtc 130  
QY 121 cacttccatgacgcgtcgtgaaatggttcgacgctacat 161  
Db 131 cacttccatgacgcgtcgtgaaatggttcgacgctacat 171

RESULT 7  
US-08-462-695-11  
; Sequence 11, Application US/08462695  
; Patent No. 5854025  
;;  
;; GENERAL INFORMATION:  
;; APPLICANT: EDWARDS, Richard Mark  
;; APPLICANT: BAWDEN, Lindsey  
;; TITLE OF INVENTION: IGF-II ANALOGUES  
;; NUMBER OF SEQUENCES: 12  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: BANNER & ALLEGRETTI, LTD.  
;; STREET: 10 S. WACKER DRIVE, SUITE 3000  
;; CITY: CHICAGO  
;; STATE: ILLINOIS  
;; COUNTRY: U.S.A.  
;; ZIP: 60606  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/462,695  
;; FILING DATE: 5-JUN-1995  
;;  
;; CLASSIFICATION: 435  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/190,029  
;; FILING DATE: 28-FEB-1994  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/GB92/01389  
;; FILING DATE: 27-JUL-1992  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: GB 9202401.7  
;; FILING DATE: 05-FEB-1992  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: GB 9116325.3  
;; FILING DATE: 29-JUL-1991  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: JOHN J. McDONNELL  
;; REGISTRATION NUMBER: 26,949  
;; REFERENCE/DOCKET NUMBER: 94,062-A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 312-715-1000  
;; TELEFAX: 312-715-1234  
;;  
;; INFORMATION FOR SEQ. ID NO: 11:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 388 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; ORGANISM: Homo sapiens  
;;  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 11..373  
;; OTHER INFORMATION: /codon\_start= 11  
;;  
;; OTHER INFORMATION: /function="gene for IGF-II/HRP fusion protein"  
;;  
;; OTHER INFORMATION: /product="Synthetic DNA"  
;;  
;; FEATURE:  
;; NAME/KEY: misc\_feature







```

1 TELECOMMUNICATION INFORMATION
2 TELEPHONE: (919)541-8614
3 TELEFAX: (919)541-8689
4 INFORMATION FOR SEQ ID NO: 22:
5 SOURCE CHARACTERISTICS:
6 LENGTH: 111 base pairs
7 TYPE: nucleic acid
8 STRANDEDNESS: single
9 TOPOLOGY: linear
10 MOLECULE TYPE: CDNA
11
12 US-08-449-315-22

```

Query Match	16.0%	Score 148.6	DB 1	Length 1131
Best Local Similarity	50.7%	Pred. No. 1.3e-40		
Matches 464	Conservative 2	Mismatches 431	Indels 18	Gaps 4

OY		4	cagtltaaccctlaactcltaacgaatlaagtcgtccaaagtgcacaatcgtgtcgac	63
Db		98	CAGTTCCTCCTTTCTCTACGCGCACACATGCCCTAGGTGGCTTGTTGTTCCAAC	157
OY		64	aacaatgltcaacgagctcagaatccgatcccagaatgcgtgtcttoaatatgctgcac	123
Db		158	GTCGTTCGCCAAGGCCTACAGACTGATGACCAGCTCTCTAAAGCATTTGCCCTTCAT	217
OY		124	tltcatgacctgttcgtgaaatggttggaagctagatatattactggaacaacacaagt	183
Db		218	TTCATATTTCCTTTGTCATATGGGTGTGAATGATCGATTCCTATTGTGTAACGTAACGGGC	277
OY		184	ttccgcactgaaaagatgatactcgggaacgctaacgcgcgcgaaggcttccaagtatc	243
Db		278	GTTATCGATAGTGAACTTAATGAGCCTCCAAATGCTGGATCCAGSAAATGACATTTGTG	337
OY		244	gatacgactgaagagctgcgcgtttagtcagaatgccacgaacagtcagttgtcgaacctg	303
Db		338	GACAACATCAAAAGCAGAAGTTGAAGAGTGTGTCACAGAGATGTTCTTCGCTGGAATC	397
OY		304	ctgactclaagctgycgaacagagcgbactctctgcagcggagcgtctctgagaatgbcg	363
Db		398	TTACCATTTTCATCTCAAATCTCGTTTTCTTGTCGGGAGSACCATAATTTGGTTTACCA	457
OY		364	ctcgtctgaaglygactlccctlaacagcatactcatalctlgccaacgcgaacttgcctt	423
Db		458	ATGGGAAGAAAAGACAGACAGAAATAGCCAATGANACT---GGAACCTCAAACTTAACCTGGT	514
OY		424	ccattcttcaaccttgcgcccgactgaagatagctttagaacagttgggtctgaatcgctg	483
Db		515	CCCCACGAAMACTTAGTGGGACTTTAAAGGCAAGTTTAAATGCAAGGCGCTTGAT---TCT	571
OY		484	agtbacctgttgtctctgcgcggagcacacactttgaaagaacccgitttaglitatc	543
Db		572	ACAGATCTCGTGGCTTATCAGGAGGCCACACGTTGTGAAATTCAMGATCCATGTTCTTC	631
OY		544	atggaatgagctcltaaaatltaogaaacacttgygttlaacttgaaccacgcgtgaacaatlag	603
Db		632	AGTGACGGCCCTCATCAACTTCAAGGCAAGGAAGACCCGACACAAAGCGTTGACCCATA	691
OY		604	tatctccaagacactgagagagctgtgtgccacgtgaatggcaacctcagltgactagtygac	663
Db		692	TACAGGGACAGCTTCGAMAGCTTTGATACATCMACAA-----CACGAGTMAAT	742
OY		664	tttatctgcgagcccccaacacatcttcgataaacagtaactatylgaatctcagaaggagcag	723
Db		743	TTTCACCCAGTCACACCCACTAGATTGTGACAAAGCCCTATTACAACAATTTGATTAGCTTA	802
OY		724	aaaagcgtbatatacagatlgatcaaagaactgtttaagcagtcagagacgcactgaacaatc	783
Db		803	AGAGGGCTTCTCCAAAGGACCAAGACTCTTCTCAACTCC---CAGAGCTGATACACA	859
OY		784	ccaactggtgagaagtlttgctcaactcaactcaaaccttcttaagcgtcttcgtggagaagc	843
Db		860	GCCATTGTCAAAACTTTTGTGCTGCCAACAGAGTGCCCTTTCTTTAAACAATTTGTGAATTC	919

[illegible]

RESULT 10  
 US-08-444-803-22  
 Sequence 22. Application US/08444803  
 Patent No. 5634414  
 GENERAL INFORMATION:  
 APPLICANT: Ryals, John A.  
 APPLICANT: Alexander, Danny C.  
 APPLICANT: Beck, James J.  
 APPLICANT: Duesing, John H.  
 APPLICANT: Friedrich, Leslie B.  
 APPLICANT: Goodman, Robert M.  
 APPLICANT: Harms, Christian  
 APPLICANT: Meins, Jr., Frederick  
 APPLICANT: Montoya, Alice  
 APPLICANT: Moyer, Mary B.  
 APPLICANT: Neuhaus, Jean-Marc  
 APPLICANT: Payne, George B.  
 APPLICANT: Sperison, Christoph  
 APPLICANT: Stinson, Jeffrey R.  
 APPLICANT: Uknes, Scott J.  
 APPLICANT: Ward, Eric R.  
 APPLICANT: Williams, Sherlica C.  
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
 NUMBER OF SEQUENCES: 106  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CIBA-GEIGY Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10532  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/444,803  
 FILING DATE: 19-MAY-1995  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/181,271  
 FILING DATE: 13-JAN-94  
 APPLICATION NUMBER: US 08/093,301  
 FILING DATE: 16-JUL-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/937,197  
 FILING DATE: 6-NOV-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/678,378  
 FILING DATE: 1-APR-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/305,566  
 FILING DATE: 6-FEB-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/165,667  
 FILING DATE: 8-MAR-1988  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/042,847  
 FILING DATE: 6-APR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/632,441  
 FILING DATE: 21-DEC-1990



```

1      FILING DATE: 24-MAY-1995
2      CLASSIFICATION: 800
3      PRIOR APPLICATION DATA:
4      APPLICATION NUMBER: 08/181,271
5      FILING DATE: 13-JAN-94
6      APPLICATION NUMBER: US 08/093,301
7      FILING DATE: 16-JUL-1993
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: US 07/937,197
10     FILING DATE: 6-NOV-1992
11     PRIOR APPLICATION DATA:
12     APPLICATION NUMBER: US 07/678,378
13     FILING DATE: 1-APR-1991
14     PRIOR APPLICATION DATA:
15     APPLICATION NUMBER: US 07/305,566
16     FILING DATE: 6-FEB-1989
17     PRIOR APPLICATION DATA:
18     APPLICATION NUMBER: US 07/165,667
19     FILING DATE: 8-MAR-1988
20     PRIOR APPLICATION DATA:
21     APPLICATION NUMBER: US 08/042,847
22     FILING DATE: 6-APR-1993
23     PRIOR APPLICATION DATA:
24     APPLICATION NUMBER: US 07/632,441
25     FILING DATE: 21-DEC-1990
26     PRIOR APPLICATION DATA:
27     APPLICATION NUMBER: US 07/425,504
28     FILING DATE: 20-OCT-1989
29     PRIOR APPLICATION DATA:
30     APPLICATION NUMBER: US 07/848,506
31     FILING DATE: 6-MAR-1992
32     PRIOR APPLICATION DATA:
33     APPLICATION NUMBER: US 07/768,122
34     FILING DATE: 27-SEP-1991
35     PRIOR APPLICATION DATA:
36     APPLICATION NUMBER: US 07/580,431
37     FILING DATE: 7-SEP-1990
38     PRIOR APPLICATION DATA:
39     APPLICATION NUMBER: US 07/368,672
40     FILING DATE: 20-JUN-1989
41     PRIOR APPLICATION DATA:
42     APPLICATION NUMBER: US 07/329,018
43     FILING DATE: 24-MAR-1989
44     PRIOR APPLICATION DATA:
45     APPLICATION NUMBER: US 08/045,957
46     FILING DATE: 12-APR-1993
47     ATTORNEY/AGENT INFORMATION:
48     NAME: Elmer, James Scott
49     REGISTRATION NUMBER: 36,129
50     REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
51     TELECOMMUNICATION INFORMATION:
52     TELEPHONE: (919)541-8614
53     TELEFAX: (919)541-8689
54     INFORMATION FOR SEQ ID NO: 22:
55     SEQUENCE CHARACTERISTICS:
56     LENGTH: 1131 base pairs
57     TYPE: nucleic acid
58     STRANDEDNESS: single
59     TOPOLOGY: linear
60     MOLECULE TYPE: cDNA
61     US-08-449-043-22

```

	Query Match	16.0%	Score 148.6;	DB 1;	Length 1131.
	Best Local Similarity	50.7%;	Pred. No. 1.3e-40;		
	Matches 464; Conservative	2;	Mismatches 431;	Indels 18;	Gaps
Oy	4	cagtaaacccatcatctcgaacaaatagctggtcccaagtgccaacatcgttgcgac	63		
Dd	98	CAGTTTTCCTCTTCTCCTACGCCACCACCATGCCCCTAGCGCTTGCTGTGGTCTTAAC	157		
Oy	64	acaactgcgaacgagtcgatccgatcccgatgcgtcttcaatatattacgtctgcac	123		

D	158	GTGGTGGCCCAAGCCCTTACAGACTGATGACCGGACCTCTGCTAAGCTCATTCGCCCTTCAT	217
O	124	ttccatgactgctcttcgtgaatggtctgcagcgtctagcatattactggaacaacaccacagt	183
D	218	TTTCATGATGTCCTTGTGCATAGGGGTGTGATGATCGATTCGATTCATTTGTGAGACGTACCGGGC	277
O	184	ttccgcactgaagaagatgcatcttcgggaacgtcaacgcgcgcaggggtcttccagtgatc	243
D	278	GTTATTCGATTGTAAGCTTTAATGAGCTTCGAAATGTGTGAATCCAAAGGAATGGAACATTTGTG	337
O	244	gacgcagtaagagctgcgcgtttgagtcagcatgcacgcacgaacagtcagtttgacagacctg	303
D	338	GACACACATCAAGAAGCAGCAGTGTGAAGAGTCCTTGTCCAGAGCTGTGTTCTTGCCCTGATATC	397
O	304	ctgaactaagctcgcgcacaagacgctgactcttcgacgcgagaccgttcctggaagatgctg	363
D	398	TTAGGCAATTTTCATCTCAAAATCTCTGTTTCTTGTGTGGGAGAGCCAAATTTGGTGTGTACA	457
O	364	ctcgctgcagctgactccctcctcagcattccctagatctgcgcacagcccaactgctgcctc	423
D	458	ATGGGAGAAAAAGCAGCAGATATGGCCAAATGAACT--GGAACCTAAACCTTAACCTGT	514
O	424	ccactcttcacccctgcgcccaagctgaagatagctcttgaacacgtggctcgtaaatcgtctg	483
D	515	CCCTCAGAAACTTATGTGTGGACTTAAAGGCAAGTTTAAAGATCAAGGGCTTGAT--TCT	571
O	484	agtgacctgtgctctgctccggagggacacacatcttggaaaagaccagtgtagtgatc	543
D	572	ACAATCTCTGGGTCTATATCAGGAGCCACACGTTTGGAAATTCAGATGCATCTTTCTT	631
O	544	atgatatgagctctacaatttcagacaacgttggtactgcacccacgcctgaacactacg	603
D	632	AGTACCGCCCTCATCTCAATTCAAGGGCAGCAGGAAGACCCGACACACAGCTTGAGCCATA	691
O	604	tatctcagacactgagaagctctgtgcgccactgaatggcgaacctcagltgactatggac	663
D	692	TACAGGGAGCGCTTTCGAGGACTTTGTACTCTCAACAAA-----CACGAGTAAAT	742
O	664	tttgatctcgagaccaccaacacatcttcgataaagaagtactgtgaaatctagaagagcag	723
D	743	TTTCAGCCCATGTCACACCCACTAGATTGTGACAAAGCCTATTTCACAACTTTGATTAGCTTA	802
O	724	aaaggcctgatatacagagtgtatcaagaactgttttagcagctccagaagcgcactgacacatc	783
D	803	AGAGGGCTTCTCCAAAGGAGCACAAGAGCTTCTCAACTCC--CAGAGCTGATACCACA	859
O	784	ccactggtgagaagttttgtcctaactctaactaaaccttcttaacgccttcgtggaaagcc	843
D	860	GCCATTGTCAAAACTTTTGGCTGCCAAGCAAGTGCCTTTCTTTAAACAATTTGTCAAAATCA	919
O	844	atggacccgtataggttaacatatacccccctcgtacggttaccacaagccagatctcgtctaac	903
D	920	ATGATCAAAATATGGGCAACCTCAACGCTCCCCCTGGCATTCGATCAAGAAGTTAGTTGGAC	979
O	904	tgcagagatggtcaac	918
D	980	TGTAAAGAGGTCAAC	994

RESULT 12  
 US-08-456-265A-22  
 : Sequence 22, Application US/08456265A  
 : Patent No. 5767369  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Alexander, Danny C.  
 : APPLICANT: Ryals, John A.  
 : APPLICANT: Goodman, Robert M.  
 : APPLICANT: Stinson, Jeffrey R.  
 : TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
 : TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
 : NUMBER OF SEQUENCES: 111  
 :  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: CIBA-GEIGY Corporation  
 :

STREET: 520 White Plains Road, P.O. Box 2005  
CITY: Tarrytown  
STATE: New York  
COUNTRY: USA  
ZIP: 10591

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US 08/456,265A  
FILING DATE: 31-MAY-95  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/181,271  
FILING DATE: 13-JAN-1994

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Weigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727/DIV10.10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8587  
TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO.: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1131 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA

US-08-456-265A-22

Query Match	16.0%;	Score 148.6;	DB 2;	Length 1131;
Best Local Similarity	50.7%;	Pred. No. 1.3e-40;		
Matches 464; Conservative	2;	Mismatches 431;	Indels 18;	Gaps 4;

QY	4	caagtaacccctatcattctacgaacaaatagctgtgccaaaglytccaacatcgtttgcgac	63
Db	98	CAGTTCCTCTCTTCTCTCTACAGCCACACACATGCCCTACGCTGCCCTTCGTGTTCAC	157
QY	64	acaaatgcacagagctcagatccgatcccaaggatcgcgtcttcataattagctcgcac	123
Db	158	GTGTTTCACATAGTGAACCTTAATGACCTCCAAATGTTGGATTCACAGGAATGACATTTG	337
QY	244	galtgcacatgaagctctgcctgttagtcagatgtcccacgacagtcagttgtcagacctg	303
Db	338	GACMACATCAAAAGACGAGATGTGAGAGTGCTGTGCCAGAGATTTGTTCTTGCTGTGATTC	397
QY	304	ctgcacatagctgcgcacacagagcgtgtacctctgtgcagcgagaccgtctctgagagctgcg	363
Db	398	TTAGCCATTTTCATCTCAAAATCTCTGTTTCTTGTCGGAGAGACCAATTTGGTTTACCA	457
QY	364	ctcgtctcagctgtacctccctacaggacltccatagatctgccaagcacaactctgcctc	423
Db	458	ATGCGAAGAAAAGACACACAGAAATAGCCAAATGAGAACT---GGAACCTCAACTTACCTGGT	514
QY	424	ccaattctcagcccgcccccgactggaagatagctttagaagaaglytggtctgaatcgtcg	483
Db	515	CCCTCAGAAACTCTAGTGGACCTTAAAGGCGMAAGTTTAAAGATCAAGGGCTGTGAT---TCT	571
QY	484	agtaacctgttgctctcgtccgcgagacacacatttgaagaagacccaglytgaatctac	543
Db	572	ACAGATCTCGTGCGCTATATCAGGAGGCCACACGCTTTGGAAATCAAGATGCATGTTCTTC	631
QY	544	atgagatagctctacaatttcacgaacaactcgtggtttacctgtgaccccaagctgtgaactag	603
Db	632	AGTGACCGCTCATCAACTTCAACGGCACAGGAAGACCCACACACAGCTGTGACCCCAATA	691
QY	604	tattctcagaacacttgaagaagctctgtgcacactgaatgacaaactcagtgacatgtggac	663
Db	692	TACAGGAGGAGCGCTTCGCAACACTTTGTACTACTTCACAAATA-----CACAGATTAAT	742
QY	664	tttgatctcgcgaaccccaacaactcttcgatatacaagtaactatgtgaatctagaagcag	723
Db	743	TTTGAGCCCACTCACACCCACTTACATTTTGACAMACACTATTNCAACAATTGTGATTACTTA	802
QY	724	aagagcctgtataagagtgtataagaagctlttagaagttccaaagcccaactgtgacacatc	783
Db	803	AGAGGGCTTTCTCAAAACGACCAAGAGAGCTTTCTCAACTCC---CAGAGCTGATPACCA	859
QY	784	ccactgtgtgaagaaglttgttaactcttaccacaactcttctaagcgcctctgtgaagcc	843
Db	860	GCAATTTGTCAAACTTTTGTGTGCACAGAACGATGCGCTTTTAAACAAATTTGTGAATCA	919
QY	844	atggaacgcatgtggtataacattaccctctgtacggtgtacccaagcgcaagatctgctgaac	903
Db	920	ATGATCAAAATGGGCAACCTCAAGAGCTCCCTCGGCAATTGCATCAAGAAAGTTAATTTGAC	979
QY	904	tgcacagatgtgtcaac	918
Db	980	TGTAAGAGGGTCAAC	994

US-08-455-46--22  
Sequence 22, Application US/08455416  
Patent No. 5777200  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Umes, Scott J.  
APPLICANT: Ward, Eric R.  
TITLE OF INVENTION: Williams, Shericca C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
NUMBER OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,416  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:

	APPLICATION NUMBER:	US 07/580,431	
	FILING DATE:	7-SEP-1990	
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER:	US 07/368,672	
	FILING DATE:	20-JUN-1989	
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER:	US 07/329,018	
	FILING DATE:	24-MAR-1989	
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER:	US 08/045,957	
	FILING DATE:	12-APR-1993	
	ATTORNEY/AGENT INFORMATION:		
	NAME:	Elmer, James Scott	
	REGISTRATION NUMBER:	36,129	
	REFERENCE/DOCKET NUMBER:	S-19825/P1/C/GC 1727	
	TELECOMMUNICATION INFORMATION:		
	TELEPHONE:	(919)541-8614	
	TELEFAX:	(919)541-8689	
	INFORMATION FOR SEQ ID NO:	22:	
	SEQUENCE CHARACTERISTICS:		
	LENGTH:	1131 base pairs	
	TYPE:	nucleic acid	
	STRANDEDNESS:	single	
	TOPOLOGY:	linear	
	MOLECULE TYPE:	CDNA	
	US-08-455-416-22		
Query Match	16.0%;	Score 148.6; DB 2;	Length 1131;
Best Local Similarity	50.7%;	Pred. No. 1.3e-40;	
Matches	464; Conservative	2; Mismatches 431;	Indels 18; Gaps
Oy	4	cagtaaacccctacattcttacgacaatgcgtgccaaagcgtgccaatcgcttcgagc	63
Dd	98	CAGTATTTCTCCTTTCCTTTCTTAGCCACCAACGATGCCCTTGACTGGTTGTGTTCTAAC	157
Oy	64	aacaatgcgaagaagcgtcagatccgatcccagaatlcgctgcttcaataltaactgcgtcac	123
Dd	158	GAGGTGGCCCAAGCCCTACAGACATGATGACCAGCTCGTCTGAAGCATTTCCGCTTCAT	217
Oy	124	tccaatgcgtccttcgtgaatagtgtttgcgaagcgtacatatatacttgacaacaccacagt	183
Dd	218	TTCATGATTTCTTTGTGCATAAAGGGTGTCATGATTCATTATGTTAGTACGTAACGGGCGC	277
Oy	184	ttccgcaactgaagaagatgcaccttcgggaagcgtacaagcgcgcaggcgcttccagtgatc	243
Dd	278	GTTATTCGATCTGAACCTTAATGCACTTCACAATGGTGGAATTCAGAAGAAATGACATTGTG	337
Oy	244	galccgaagaagggcgtccgtttgaatcagaatgccacgaacagtcaglttgtcagacctg	303
Dd	338	GACAACATCAAAGCAGCAGTGTGAGAGTCTTGTCCAGAGATGTTGTTCTTGCGCTGATATC	397
Oy	304	ctgactatagctgcgcaacagaagcgtgcactctgcaggcgagccgttccttgagaagtgcg	363
Dd	398	TTAGGCATTTCATCCAATACTCTGTGTTTTTGTGCGGAGGACCAATTTGGGTTGTAACA	457
Oy	364	ctcgtgcagctgtagatccccctcaggaatcctctagatctgycaccaagccaacttcgtgct	422
Dd	458	ATGGGAACAAAAAAGACAGCAGATATGCCAAATAGAACT--GCAACCTCAAAATTNACCTGGT	514
Oy	424	ccatctcttcaaccoccccgagctgaagaatgactttaagaacgctgggtlccgaatcgtcg	488
Dd	515	CCCCACGAAACTCTTAGTGGGCACTTAAAGCAAGTTTAAGATCAAGGCGCTTGAT--TCT	571
Oy	484	agtgaccttgcgtctcgtccgaggaacacalcitlgaaagaacagcgtlgaugtlcatc	543
Dd	572	ACAGATCTCGTGCTCTCTATCAGGAGCCACACGTTGGAAAATCAAGATGCATGTTCTTC	631
Oy	544	atgataggtcctctacaatttagcaaaacacgtggttactctgaagcccaagctggaacatacg	603
Dd	632	AAGTACCCCCTTCATCAACTTTCATACGCAACAGCAAGACCCGCAACAGCTTGAACCCATA	691
Oy	604	tatctcccaagacacgaagggcgtttgcccacatgaatggcaacccctcagatgcactgaatgac	663

Query Match	16.0%;	Score 148.6;	DB 2;	Length 1131;
Best Local Similarity	50.7%;	Pred. No. 1.3e-40;		
Matches 464;	Conservative 2;	Mismatches 431;	Indels 18;	Gaps 4;

Db 692 TACAGGAGCAGCTTCGACAGCTTGTACTACTACAA-----CACGAGTAAT 742  
QY 664 ttgatctggagcccaaccatctcgataaagaagactgtaatcagaaggagcag 723  
Db 743 TTGAGCCACTACACCCACTAGATTGACAGACCTATTACAACTTGTGATTAGCTTA 802  
QY 724 aaagccctgatacagaagtgaacaagactgttagcagtcacagcgaactgacaccatc 783  
Db 803 AGAGGGCTTCTCCAAAGCGACCAAGAGCTCTTCTCAGACTCC---CAGAGCTGATACACCA 859  
QY 784 ccactggtagaagtttctgaactctacccaactcttcttaagccttcgtagaagcc 843  
Db 860 GCCATTGTCTCAAACTTGTCTGCCAAGACAGTGCCTTTTAAACAATTGTGCAAAATCA 919  
QY 844 atgaccctgtagtaaccctctgacggtacccaagccagatctctgtaac 903  
Db 920 ATGATCAAAATGGGCAACCTCAAGCCTCCCTGTGGCAFTTGATCAGAGATTAGTGGAC 979  
QY 904 tgcagaagtgtcaac 918  
Db 980 TGTAGAGGCTCAAC 994

RESULT 14  
US-08-455-244-22  
Sequence 22, Application US/08455244  
Patent No. 5789214  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericea C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,244  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/Pl/CCG 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1131 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-455-244-22

Query Match 16.0%; Score 148.6; DB 2; Length 1131;  
Best Local Similarity 50.7%; Pred. No. 1.3e-40;  
Matches 464; Conservative 2; Mismatches 431; Indels 18; Gaps 4;

QY 4 caattaccctctacattcttaacgaatagatgctcccaagctgccaacatcgttcgagc 63  
Db 98 CAGTTCCTTCCTTCTTCTACGCCACACATGCTCCAGCTGCTTTCGTTGTTCTCAAC 157  
QY 64 acaatcgtcaacgagctcagatccgaatcccaagatcgtctgctcaatlatcagctgcac 123  
Db 158 GTGGTTGCCAACCCCTACAGACTGATGAGCCGAGCGTGTCTAAGCTCATTCGCCCTCCAT 217  
QY 124 ttccatgactgctcgtagatggtcgcagcgtagcatattactgacaacaccacagat 183  
Db 218 TTTCATGATGCTTGTGTCATGAGGTGTGATGATTCATTCTATTGGTAGACGTCGCGGC 277  
QY 184 ttccgacgtgaaagatgcatcgcgggaacgctaacaagcgccaggggcttccagtgatc 243  
Db 278 GTTATGATGATGACTTAATGAGCTCCAAATGTGGAATCCAGGAATGACATTTGG 337







---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 05:46:18 ; Search time 1168.17 Seconds  
(without alignments)  
3499.414 Million cell updates/sec

Title: US-09-246-451-16

Perfect score: 927  
Sequence: 1 atgcagtaacctacatc.....gagtgtaacagcaactct 927

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*

44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: em\_est20:\*  
50: em\_est21:\*  
51: em\_est22:\*  
52: gb\_est30:\*  
53: gb\_est31:\*  
54: gb\_est32:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: em\_est35:\*  
83: em\_est36:\*  
84: em\_est37:\*  
85: gb\_est48:\*  
86: gb\_est49:\*  
87: gb\_est50:\*  
88: gb\_est51:\*  
89: gb\_est52:\*  
90: gb\_est53:\*  
91: gb\_est54:\*  
92: gb\_est55:\*  
93: gb\_gss1:\*  
94: gb\_gss2:\*  
95: gb\_gss3:\*  
96: gb\_gss4:\*  
97: em\_gss1:\*  
98: em\_gss2:\*  
99: em\_gss3:\*  
100: em\_gss4:\*  
101: gb\_gss5:\*  
102: gb\_gss6:\*  
103: em\_gss5:\*  
104: em\_gss6:\*  
105: gb\_gss7:\*  
106: gb\_gss8:\*  
107: gb\_gss9:\*  
108: em\_gss7:\*  
109: em\_gss8:\*  
110: gb\_gss11:\*  
111: gb\_gss10:\*  
112: em\_gss9:\*  
113: em\_gss10:\*  
114: em\_gss11:\*  
115: em\_gss12:\*  
116: gb\_gss12:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	290	31.3	620	62	AV441260	AV441260 AV441260
2	276.6	29.8	516	62	AV442242	AV442242 AV442242
C 3	236.4	25.5	720	122	CNS00WMO	AL094170 Arabidops
4	234.8	25.3	652	118	A0956990	A0956990 LERN24TR
5	210	22.7	564	20	AA042293	AA042293 24630 CD4
6	204.6	22.1	537	29	AA650756	AA650756 30870 Lam
7	204.4	22.0	824	46	AI959837	AI959837 sc94h07.y
8	187	20.2	499	90	T88184	T88184 11880 Lambd
9	185.4	20.0	454	62	AV441600	AV441600 AV441600
10	184.2	19.9	475	87	N37495	N37495 18722 Lambd
11	176.2	19.0	461	85	H37375	H37375 15504 Lambd
12	172.4	18.6	635	72	AA24453	AA24453 707014P11
13	172.2	18.6	635	72	AA278775	AA278775 sf97d02.y
14	169.6	18.3	745	74	AW559660	AW559660 EST314772
C 15	164.4	17.7	360	89	T03969	T03969 172 Ar-NHC
16	161.6	17.4	387	62	AV439974	AV439974 AV439974
17	159.8	17.2	688	74	AW585492	AW585492 EST317115
18	158.4	17.1	497	25	AA394440	AA394440 26037 Lam
19	154.4	16.7	308	89	T45859	T45859 9122 Lambda
20	154	16.6	485	89	T42100	T42100 5363 Lambda
21	153.8	16.6	668	74	AW584286	AW584286 N210287e
22	153.6	16.6	638	74	AW587236	AW587236 EST318859
23	151.6	16.4	672	69	AW185769	AW185769 se59d08.y
24	151.2	16.3	321	85	F15487	F15487 ATTS5735 Or
25	150	16.2	592	80	AW704659	AW704659 SK54h10.y
26	149.8	16.2	640	74	AW584357	AW584357 N210383e
27	149.2	16.1	295	92	Z29133	Z29133 ATTS2079 Ve
28	148	16.0	602	80	AW705617	AW705617 SK50d03.y
29	147.8	15.9	479	89	T42049	T42049 5312 Lambda
30	147.8	15.9	658	79	AW685235	AW685235 NF027H10N
31	147.6	15.9	584	80	AW705730	AW705730 SK51b02.y
32	146.4	15.8	607	44	AI782928	AI782928 614006F01
33	146.2	15.8	632	74	AW584356	AW584356 N210382e
34	145.2	15.7	639	70	AW256487	AW256487 EST304624
35	145.2	15.7	646	70	AW267813	AW267813 EST305941
36	144	15.5	475	89	T22196	T22196 4204 Lambda
C 37	144	15.5	527	118	A0965253	A0965253 LER1B21TR
38	142.8	15.4	659	74	AW559945	AW559945 EST314993
39	139.4	15.0	608	80	AW705946	AW705946 SK52h07.y
40	138.4	14.9	668	79	AW685437	AW685437 NF029D09N
41	136.6	14.7	583	46	AI960823	AI960823 sc91a06.y
42	136.4	14.7	670	74	AW574244	AW574244 EST316835
43	136.4	14.7	690	74	AW586142	AW586142 EST317765
44	135.8	14.6	377	30	AA728602	AA728602 33011 Lam
45	135.4	14.6	703	74	AW561032	AW561032 EST316080

ALIGNMENTS

RESULT 1  
AV441260/c  
LOCUS  
DEFINITION  
AV441260 620 bp mRNA  
EST  
19-APR-2000  
AV441260 Arabidopsis thaliana above-ground organ two to six-week  
Old Arabidopsis thaliana cDNA clone Ap237e04\_f 3', mRNA sequence.

ACCESSION	AV441260	GI:7611644
VERSION	AV441260.1	
KEYWORDS	EST.	
SOURCE	thale cress.	
ORGANISM	Arabidopsis thaliana	
REFERENCE	Nakamura, Y. and Asamizu, E.	
AUTHORS	Arabidopsis thaliana expressed sequence tags	
TITLE	Unpublished (2000)	
JOURNAL	On Jul 7, 1999 this sequence version replaced gi:5410020.	
COMMENT	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/. Localization/Qualifiers	
FEATURES	1. 620	
source	/organism="Arabidopsis thaliana"	
	/strain="Columbia"	
	/db_xref="taxon:3702"	
	/clone="Ap237e04_f"	
	/clone_lib="Arabidopsis thaliana above-ground organ two to six-week old"	
	/issue_type="above-ground organ"	
	/dev_stage="two to six-week old"	
BASE COUNT	178 a 113 c 147 g 182 t	
ORIGIN		
Query Match	31.3%; Score 290; DB 62; Length 620;	
Best Local Similarity	Pred No. 4.2e-78;	
Matches 377; Conservative 0; Mismatches 145; Indels 0; Gaps 0;		
QY 406	aacgcacactgctcgtccatcttcacccgagtcgaggaatagcttagaac	465
DB 620	AATGCAATCTTCACCTCATTACCTTACCTTCAAGCAGCTTACAGAAAT	561
QY 466	gtgggtcgtgaatcgtccgagtgacctgtgctcgtccgaggaacacatttgaaag	525
DB 560	GTGGGCTCATGCTCTTCTGATCTCGTGTGCTCCGCTGATACACATTTGGTAA	501
QY 526	aaccagtgatgtcctcaatgaatgagcttacaatttgaagcaactggttacctgac	585
DB 500	AATCAATGTCAGTTTATTTCTTGACAGATTATACAAATTTGACACAGGTTTACCGAC	441
QY 586	cccacgctgacacactacgtatctccagacactlgagagcttggccactgaatggcaac	645
DB 440	CCYACACTCAACCTTACTTACCTTCAAGCTCTCGTGAGCTATGCCCTTATGGCAAT	381
QY 646	ctgaatgacactagtgactttagcttcgctggaaccccaacatcttctgataaagtaactat	705
DB 380	CGAAGTGCCTTGGTAAATTTGATCTACGATACGCTTACGCTTTCACACAAATTAATCTAC	321
QY 706	gttaactctgagaggaagaagcccgatatacaagatgataaactgttttagcaatca	765
DB 320	GTTAACTCTCAAGAGCGAAGAAAGTCTTATTCAGAGCGACCAAGAGTTGTTCTTACCC	261
QY 766	gaagcactgacacacatcccaactggtgagagaagtttgcctaaacttaactaaacactctt	825
DB 260	AATGCCACATGACACAAATCCCTTGGTGAAGCATATGCGATATGGACACACAAATCTTC	201
QY 826	aaagccttgatgaaagcactgaacgtaaggtaacattaccctctgacggtaccca	885
DB 200	AATGCAATTTGTGAGGCAATGATAGATGGGAAACATTACCACTTACAGAACTCA	141
QY 886	ggccagatcgtctgaactgacgacgagtgatcaagaagaactct 927	
DB 140	GGACAAATCAGATTGAACTGTAGAGTTGTGAATCTCAACTCT 99	

LOCUS	2	AV442242	516 bp	mRNA	EST	19-APR-2000
DEFINITION		AV442242	Arabidopsis thaliana above-ground organ two to six-week old Arabidopsis thaliana cDNA clone AP237604_r 5', mRNA sequence.			
ACCESSION		AV442242				
VERSION		AV442242.1	GI:7612647			
KEYWORDS		EST.				
SOURCE		thale cress.				
ORGANISM		Arabidopsis thaliana				
REFERENCE		Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS		1 (bases 1 to 516)				
TITLE		Nakamura,Y. and Asamizu,E.				
JOURNAL		Arabidopsis thaliana expressed sequence tags unpublished (2000)				
COMMENT		Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yama 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/. Location/Qualifiers				
FEATURES		source				
		1..516				
		/organism="Arabidopsis thaliana"				
		/strain="Columbia"				
		/db_xref="taxon:3702"				
		/clone="Ap237604_r"				
		/clone.lib="Arabidopsis thaliana above-ground organ two to six-week old"				
		/issue-type="above-ground organ"				
		/dev-stage="two to six-week old"				
BASE COUNT		138 a 129 c 107 g 142 t				
ORIGIN						
Query Match		29.8%; Score 276.6; DB 62; Length 516;				
Best Local Similarity		71.1%; Pred. No. 5e-74;				
Matches 366; Conservative		1%; Mismatches 149; Indels 0; Gaps 0;				
169	gacacacaccacaggttcgcgcacgtaaagaatgcatcggaagcgtacacgcgcagc	228				
2	GACAAACAGACATCTCTTTCGAAACAGAAAGATGCGTTGGAAAGCAATTCGGCTGG	61				
229	ggcttcagatgatcatgcacgcacgaagcgtccggttgatgcagcctgcacgaacagtc	288				
62	GGATTTCAGATGATGATGAATGAAGTGAACGTGGGGAGAGGCGATGCCCAAGAACCGTT	121				
289	agctgtcagacgtcgtcactatagctcgcgaacagagcgtaactcttgcaagcgagaccg	348				
122	TCATTCGCGAGATATCTCACCATTCACACTTCACATATGTCACATTCGGCAGAGGTCCT	181				
349	tccttgagagagcgcgctcgtgtcgaagtgtaactccctacacggatcttcttgatctggcaac	408				
182	TCCTGGAGGGTTCCTTTGGGAAAGAGACACATTTACAAGCATTCCTGGAATCTCCTAT	241				
409	gccaacctgcctgcctccattcttcacccctgcgccacgtgaagatagacttgaaacgtg	468				
242	GCAATTCCTCCAGCTCCATCTTTTACACTTCACACAACCTTAAAGCCAGCTTCAGAAATGTT	301				
469	ggtctgaatcgcctcgaatgacctgttgacctctgtccgagagacacacatttgaaagaac	528				
302	GGTCTCGATTCGCTCCCTTCGATCTCGTTCCTCTCCGGGTCACACATTTTGGTAAAT	361				
529	cagtttagatcatcatatgtagtaggtctcaaatcttcagcaaacctgggttaactgaccc	588				
362	CAATCTCAGTTTATCTTTCGACAGATTATACAAATTTTCAGCAACACAGGTTTACCCACCT	421				
589	agcgcgaacactatcttcacgaacactggaagcgtctgtgcacactgaatgcaacctc	648				
422	ACACGCAACACACTACTACCTCCAAACTCTGCTGTGACATATGCCCTTAATGCAATGCA	481				
649	agtgcaactagtgaccttgatctctgcggaacccaac	683				

Db	482	AGTGGCTTGTAAGATTTCATCTACGTACGCGCTAC	516
RESULT	3		
CNS00MXO/c			
LOCUS			
DEFINITION	CNS00MXO 720 bp DNA	GSS	28-JUN-1999
ACCESSION	Arabidopsis thaliana genome survey sequence T7 end of BAC T12L18 of TMMU library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.		
VERSION	AL094170		
KEYWORDS	AL094170.1 GI:5295324		
SOURCE	GSS.		
ORGANISM	thale cress.		
REFERENCE	Arabidopsis thaliana		
AUTHORS	Euhayriota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 1 to 720)		
AUTHORS	Salamonbat,M., Choisine,N., Attiguenave,F., Brottier,P., Wincker,P., Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.		
JOURNAL	unpublished		
REFERENCE	2 (bases 1 to 720)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
FEATURES	Location/Qualifiers		
source	1..720		
	/organism="Arabidopsis thaliana"		
	/strain="Columbia"		
	/db_xref="taxon:3702"		
	/clone_id="TMMU"		
	/clone="T12L18"		
	/note="end : T7"		
BASE COUNT	229 a 146 c 134 g 211 t		
ORIGIN			
Query Match	25.5%; Score 236.4; DB 122; Length 720;		
Best Local Similarity	71.9%; Pred. No. 1.2e-61;		
Matches	309; Conservative 0; Mismatches 121; Indels 0; Gaps 0;		
07	438 tctgtccgagagacacacatttggagaagaaccagtggtatcatcatgatagagctcta	557	
Db	644 TTTGTTAGGGGCTCACATTGTTGTAATAATCAATGATGCATTATGAGACAGATTATA	585	
07	558 caattcaggcaacatcgtgttaccctgaccccccgcgttgaaactaagttctccagaact	617	
Db	564 CAACCTTTTGACACACTGTGATTACCTGACCCTTACACTCAACACTACTTACCTCCAAACCT	525	
07	618 gagagagcttggccccacgtgaacactcaagtcagtcacatgagagcttgaatcgtcgag	677	
Db	524 TCGGTGTCATATGCTCTCGCAATGATGTAATCAAAAGCGTTTGAAGTTCGATCTCGTAC	465	
07	678 ccccaacatctcgtataaagaagtaactatgtaacttgaagaagaagaagcctgataca	737	
Db	464 GCCTTTGGTTTGGCACAACAATACTATGATGTAATTTAAAGCGCAAAAAGGCTTATCCA	405	
07	738 gagtgatcaagaacggtttagcagtcagaagccacactgacacccacccacgtgtagaag	797	
Db	404 GAGCGACCAAGAGTGTCTCTAGGCCCAATGCCACTGACACAAATCCCTTGGTGATGAGC	345	
07	798 ttgttaacttactcaaacctctttaaagcgtctgtgaaagcaltgagaccgtatagg	857	
Db	344 ATATGCTGATGCGACACAAACATTTCTTCAATGATCATGCTGGAGCAATGAATAGATGG	285	
07	858 taagctacccctcgtgaaggtgaccagaagccaggtctgtctgaactcagatggtgtaa	917	
Db	264 AATATTTACCACTACACAGAACTCAAGAGCAAAATTCAGGTTGATTTGAGGTGTAA	225	

QY	918	cagcaacct	927	1	
Db	224	CTCCACTCT	215		
RESULT	4				
LOCUS	A0956990				
DEFINITION	A0956990	652 bp	DNA	GSS	28-JAN-2000
ACCESSION	A0956990				
VERSION	A0956990.1	GI:6784691			
KEYWORDS	GSS.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota, Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta				
AUTHORS	Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;				
	Brassicaceae; Arabidopsis.				
	1 (bases 1 to 652)				
	Buell, C.R., Lin, X., Pal, G., Barnstead, M., Bowman, C., Utterbach, T.,				
	Feldblyum, T., Liang, F., Creasy, T. and Fraser, C.M.				
	Genomic survey sequencing of Landsberg erecta ecotype of				
	Arabidopsis thaliana and identification of sequence-based				
	polymorphisms				
	Unpublished (2000)				
JOURNAL	On Dec 15, 1999 this sequence version replaced gi:4575946.				
COMMENT	Contact: Xiaoying Lin				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0200				
	Fax: 301 838 0208				
	Email: ateligr.org				
	For additional information, see <a href="http://www.tigr.org/tdb/at/at.html">http://www.tigr.org/tdb/at/at.html</a>				
	Seq primer: TR				
	Class: shotgun.				
FEATURES	Location/Qualifiers				
SOURCE	1..652				
	/organism="Arabidopsis thaliana"				
	/strain="Landsberg erecta"				
	/db_xref="taxon:3702"				
	/clone="LERAN24"				
	/clone_lib="LERAN"				
	/note="Organ: Leaf; Vector: pROS1. Total genomic DNA was				
	sheared to 0.9-1 Kbp before ligation."				
BASE COUNT	190 a 145 c 121 g 196 t				
ORIGIN					

QY	798	ttttgcttacccttactcaaacctctttaaagcctcttgtaagccatgagccatgagg	857
Db	488	ATATGCTGATGGCACACAAACATTTCTTCAATATGCATTCCTGGAGGCAATGAATGAGATGGG	547
QY	858	taacattaacccctctgaagggataccacaagccagaatctgcgtctgaactcagaatggtcaa	917
Db	548	AAATATATACACCAACTACAGGAACTCAAGGACCAATTCAGTGTGAATGTAGAGTGCTGAA	607
QY	918	cagcaactct 927	
Db	608	CTCCAACTCT 617	
RESULT	5		
LOCUS	AA042293	564 bp	mRNA EST 19-SEP-1997
DEFINITION	24630 CD4-13 Arabidopsis thaliana cDNA clone E6A9T7, mRNA sequence.		
ACCESSION	AA042293		
VERSION	AA042293.1	GI:2414082	
KEYWORDS	EST.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
	Euaralyota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 1 to 564)		
AUTHORS	Newman,T., deBrujin,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Ralkehl,N., Somerville,S., Thomasnow,M., Retzel,E. and Somerville,C.		
TITLE	Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones		
JOURNAL MEDLINE	Plant Physiol. 106, 1241-1255 (1994)		
COMMENT	95148729		
	On Sep 19, 1997 this sequence version replaced gi:1520451.		
	Contact: Thomas Newman		
	MSU-DOE Plant Research Laboratory		
	Michigan State University		
	MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI		
	Tel: 517-353-0854		
	Fax: 517-353-9168		
	Email: 22313tcn@lhm.cl.msu.edu		
	Seq primer: 77.		
FEATURES	Location/Qualifiers		
source	1..564		

Query Match	25.3%	Score 234.8	DB 118	Length 652
Best Local Similarity	71.6%	Pred. No. 3,66-61		
Matches 308	Conservative 0	Mismatches 122	Indels 0	Gaps 0
QY	438	tctgtcggagagacacacatttggaaagaacacagtgtatgtatcatalcatalgtatagctcta	557	
Db	188	ttttgttagggcctccacacattttggtaaaaaaacaaatgtgatctattatggacagatttata	247	
QY	558	caatttcgcaacactagggttaacctgaccccaagctgtgaacctagcatcttcacacact	617	
Db	248	caacttttagcaaacactgtgatttaccctggacccttaccactcaaacactacttccccaacctt	307	
QY	618	gagagctgtgtgccacatgaaatgtgaacctagtcacatagtgagacttgaatctgtcgac	677	
Db	308	tctgtgtcaatgtctccgcataatgtaaacaaagcgttttagtgatttcgatctccgtac	367	
QY	678	cccaacacatcttgtataacaagtactatgtgaatctatagagagagcaaaagcctgtatca	737	
Db	368	gcccttttggtttcgacacaaatctatgtgaatttttaaaagccttaattatcca	427	
QY	738	gagtgatcaagaactggttttagcaatctccagacacactgacacaccatccactgtatagaag	797	
Db	428	gagcgaccacaagattgtttctcttagccccaatggccactacacacaatcccttggtagagac	487	

/note="Vector: plusscript SK-; Site-1: EcoRI; Site-2: EcoRI; using 5 ug of polyadenylated mRNA from 3 day-old *Arabidopsis thaliana* (Columbia) seedling hypocotyls as template and oligo d(T) as primer, first strand synthesis was catalyzed by Moloney murine leukemia virus reverse transcriptase (Pharmacia). Second-strand cDNA was made using the procedure of Gubler and Hoffman (1983) except that DNA ligase was omitted. After the second strand reaction, the ends of the cDNA were made blunt with Klenow fragment and EcoRI/NotI adaptors (Pharmacia) were ligated to each end. The cDNA was purified from unligated adaptors by spun-column chromatography using sephacryl s-300 and size-fractionated on a 18 low melting point mini-gel. Size selected cDNAs (0.5 - 1 kb) were removed from the gel using agarase (New England Biolabs), phenol:chloroform extracted and precipitated using 0.3 M NaOAc (pH 7)/ethanol. A portion of each cDNA size-fraction (0.1 ug) was co-precipitated with 1 ug of lambda-dazapiI (Stratagene/EcoRI digested, dephosphorylated arms and then ligated in a volume of 4 ul overnight. Each ligation mix was packaged in vitro using GigaPack II gold





Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

# REFERENCE

1 (bases 1 to 824)

# AUTHORS

Shoemaker, R., Keim, P., Vodka, L., Erpelting, J., Coryell, V., Khanna, R., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Thelsting, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterson, R. and Wilson, R.

# TITLE

Public Soybean EST Project

# JOURNAL

Unpublished (1999)

# COMMENT

On Jan 19, 1998 this sequence version replaced gi:2284894.  
Contact: Shoemaker R./Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or Info@genomesystems.com web site: www.genomesystems.com  
Possible reversed clone: similarity on wrong strand  
High quality sequence stop: 424.

# FEATURES

source

1. 824  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1019-1046"  
/clone\_lib="Gm-c1019"  
/issue\_type="Immature seed coats of greenhouse grown plants"  
/lab\_host="DH10B (Gibco BRL)"  
/note="Vector: pSPORT1 (Life Technologies); Site\_1: Not I; Site\_2: Sal I; This cDNA library was constructed from mRNA isolated from Immature seed coats (200-300 mgs) of greenhouse grown plants. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E. coli Electromax DH10B host cells (Gibco BRL). This library was constructed by Dr. Lila Vodka and Dr. Anu Khanna."

# BASE COUNT

220 a 198 c 165 g 236 t 5 others

# Query Match

22.0%; Score 204.4; DB 46; Length 824;

# Best Local Similarity

55.3%; Pred. No. 8.1e-52;

# Matches

413; Conservative 0; Mismatches 331; Indels 3; Gaps 1;

QY 4 caattaccctcatctacgacgaatagctgtcccaagctgcacacactcgttcgagac 63  
DB 81 CACCTTACTCTCTAGTCTTCAAGAAACATGTCCAAATGTGTCCTATTGTGTTGGA 140  
QY 64 acaatcgtaacagagctcagatccgatacagatcgtctcgaatattacgtctcag 123  
DB 141 GATATCTTCGATGCTTCTTTCACCGAATCCCGAATCGGSCAGTCTCATAGAGCTTCAT 200  
QY 124 ttcacatgactgctcgtgaatggttcgacgtagacalactactgacacaccacagt 183  
DB 201 TTTTCATGATGATGCTTTTCAAGCTTGATGATGATCTTTGCTGACACACATGATACA 260  
QY 184 ttcgcacatgaagaagatgattcgcggaacgctaacagcgcgcgggcttcacgagatc 243  
DB 261 ATAGAAAGCGAGAGATGACATTCACAAATATCAACTCAATTAAGAGATTGAGAGCTTGC 320

QY 244 gatccatgaaggctgcgtctgagtcagcatgcccaacagctcgttcgagacctg 303  
DB 321 AATGACATCAAGACAGCGGGTGAAGAAATAGTTGTCCACACACAGTTTCTTGTGATATT 380  
QY 304 ctgactatagctgcgaacacagagctgactctgcagcgagacctcgttcgagagtcgcg 363  
DB 381 CTTGCTTTTCAACTCACTCACTCACTTAAAGCTTCTTCTTCTTCAAGTCTCAA---CACC 557  
QY 484 agtgaacctgtgctcgttcgcggaagacacacattggaagaagacagtgatgctc 543  
DB 558 CTTGATTTAGTTACACTCTCANGTGTGATAGCGTNGAAGAGCTCGTGATACATTTC 617  
QY 544 atgagatagctcgaacttcacacacacacagctggttaacctcgaacacagctgagac 603  
DB 618 ATANACGATTATACACTTACACTTACACTGGAACCTGATCCACTGGAACACTTACA 677  
QY 604 tatcccaagacactgagagagctgtgcccactgaatgcaacactcagtgactagtgagac 663  
DB 678 TACTTTAGATTATTTGCGTGTAGATGCCCCAGAAATGGCAGTGTATNNCCACCAAT 737  
QY 664 ttgatctgcgagaccccaacacacttcgatalaacaagtaactatgtgaactagagagcag 723  
DB 738 NTGAGACTGTGACACACTGTATCAATTTGACACAGATACTACTCCATCTTCTGACGCTT 797  
QY 724 aaagcctgtacagagtgatgaagaa 750  
DB 798 AATGGCTACTCTGAGTGTGACAAAGAA 824

# RESULT

8 T88184 499 bp mRNA EST 09-JAN-1998

# LOCUS

T88184 11880 Lambda-PRL2 Arabidopsis thaliana cDNA clone 157E227, mRNA

# DEFINITION

sequence.

# ACCESSION

T88184

# VERSION

T88184.1 GI:36039

# KEYWORDS

EST.

# SOURCE

thale cress.

# ORGANISM

Arabidopsis thaliana

# REFERENCE

1 (bases 1 to 499)

# AUTHORS

Newman, T., dehrn, F.J., Green, P., Keegstra, K., Kende, H.,

# TITLE

McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas, M.,

# JOURNAL

Genes galore: a summary of methods for accessing results from

# MEDLINE

large-scale partial sequencing of anonymous Arabidopsis cDNA clones

# COMMENT

On Apr 14, 1993 this sequence version replaced gi:716536.

# CONTACT

Contact: Thomas Newman

# MSU-DOE Plant Research Laboratory

Michigan State University

# MSU-DOE-PRU, Michigan State University, Plant Biology Bldg., E.

Lansing, MI

# Tel: 517-353-0854

Fax: 517-353-9168

# Email: 22313tcn@ldm.cl.msu.edu

Seq primer: T7 dye primer.

# Location/Qualifiers

1.499

# FEATURES

Location/Qualifiers

# source

/organism="Arabidopsis thaliana"

# /strain="var columbia"



MSU-DOE Plant Research Laboratory  
Michigan State University  
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.  
Lansing, MI  
Tel: 517-353-0854  
Fax: 517-353-9168  
Email: 22313tc@ldm.cl.msu.edu  
Seq primer: T7 dye primer.  
Location/Qualifiers

## FEATURES

source

1..475

/organism="Arabidopsis thaliana"

/strain="var columbiana"

/db\_xref="taxon:3702"

/clone="213L167"

/clone\_lib="Lambda-PRL2"

/note="Vector: lambda Zip-lox; Site.1: Sal; Site.2: Not; lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRL's lambda Zip-lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dr primed cDNA."

BASE COUNT 124 a 112 c 94 g 126 t 19 others

ORIGIN

Query Match

Best Local Similarity 71.4%; Score 184.2; DB 87; Length 475;  
Pred. No. 1e-45;

Matches 240; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

OY 560 attcagcaacactggtgctacacgccccgcgtgaacactagctatctccagacactga 619  
|||||  
DB 1 ATTTCAGCAACAGGTTTACCGACCTTACTCACTCACTACTTCTCCAAACTCTTC 60  
  
OY 620 gaagctctgccccactgaatggaacactcagtcagctagtgagacttgatctgcgagcc 679  
|||||  
DB 61 GTGAGACTATGCCCTTAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 120  
  
OY 680 caaccatctcgataaagaactatgtaactctagaagagagaagagagagagagagagag 739  
|||||  
DB 121 CTACGGTTTTCGACAACTTACTAGCTCAATCTCAAGGCAAGGCAAGGCTTATCTCAGA 180  
  
OY 740 gataaagaactggttagaagtcagacgcacacacacacacacacacacacacacacacac 799  
|||||  
DB 181 GCGACCAAGAGTTGTTCTCTAGCCCAATGCCACTGACACAAATGCCCTTGTGAGAGCAT 240  
  
OY 800 ttgctaactactcaaacctctcttaacgccttcgtggaagagcagtcagtcagtcagtc 859  
|||||  
DB 241 ATGCTGATGCGACACAAACTTCTTCATATGCAATTTGGAGGCAATGATGATGGGAA 300  
  
OY 860 acattacccctctgacgggtgatacccaagagcagattc 895  
|||||  
DB 301 ACATTACACCACTACAGGAGCCTCAAGNCANATC 336

RESULT 11

LOCUS

H37375

DEFINITION

H37375

ACCESSION

H37375.1

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE

1 (bases 1 to 461)

AUTHORS

Newman,T., deBrujin,F.J., Green,P., Keegstra,K., Kende,H.,

McIntosh,L., Ohlroge,J., Raikhel,N., Somerville,S., Thomasow,M.,  
Reisel,E. and Somerville,C.  
Genes galore: a summary of methods for accessing results from  
large-scale partial sequencing of anonymous Arabidopsis cDNA clones  
Plant Physiol. 106, 1241-1255 (1994)  
95148729

COMMENT

On Dec 30, 1997 this sequence version replaced gi:906874.

CONTACT

Contact: Thomas Newman

MSU-DOE Plant Research Laboratory

Michigan State University

MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.

Lansing, MI

Tel: 517-353-0854

Fax: 517-353-9168

Email: 22313tc@ldm.cl.msu.edu

Seq primer: T7 dye primer.

Location/Qualifiers

1..461

/organism="Arabidopsis thaliana"

/strain="var columbiana"

/db\_xref="taxon:3702"

/clone="180H777"

/clone\_lib="Lambda-PRL2"

/note="Vector: lambda Zip-lox; Site.1: Sal; Site.2: Not; lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRL's lambda Zip-lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dr primed cDNA."

BASE COUNT 120 a 103 c 85 g 135 t 18 others

ORIGIN

Query Match

Best Local Similarity 71.1%; Score 176.2; DB 85; Length 461;  
Pred. No. 2.9e-43;

Matches 243; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

OY 424 coactctcaaccctcccaagctgaagatagcttagaagcgtggtcgaacgcgcg 483  
|||||  
DB 1 CCATTCTTACNCTCCACACACTTAAGGCCAGTTTAAATGTTGACTTGCACCTGCC 60  
  
OY 484 aagcactctgctctgctccggaagacacactttggaagaagcagtgatgctc 543  
|||||  
DB 61 TCTNATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
  
OY 544 atgataagctctacaattcaacaacactggttaccctgaccccaagctgaacactagc 603  
|||||  
DB 121 ATGATATAGACTATACAACTTATACACACTGTTTACCGACCTTACTTATATAC 180  
  
OY 604 tatctcagaacactgagagctgtgcccactgaatgcgaacctcagtcagtcagtcagtc 663  
|||||  
DB 181 TACCTCAGACACTTGTGTGACAACTGTCCTCCCGCAATGATGATGATGATGATGATGAT 240  
  
OY 664 ttgatactcggaagcccaacacactctcgataaagaactatgatacttagagagcag 723  
|||||  
DB 241 TTCGATCTACGACACACGAGCGGTTTGTGACAAATAAATACTATGTAATGTAAGAGCTA 300  
  
OY 724 aaagcctgatacaga-gtcatcaagaactgtttagcagtc 764  
|||||  
DB 301 AAGGGCTTATCCAGACGGGATCAAGAGTTGTTTCCAGCC 342

RESULT 12

LOCUS

AM424453

DEFINITION

AM424453

ACCESSION

AM424453.1

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE

1 (bases 1 to 625)

AUTHORS

Newman,T., deBrujin,F.J., Green,P., Keegstra,K., Kende,H.,

KEYWORDS	EST
SOURCE	Zea mays.
ORGANISM	Zea mays.
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
AUTHORS	1 (bases 1 to 625)
TITLE	Walbot.V.
JOURNAL	Maize ESTs from various cDNA libraries sequenced at Stanford University
COMMENT	Unpublished (1999)
COMMENT	On May 7, 1998 this sequence version replaced gi:3119571.
FEATURES	Contact: Walbot V
source	Department of Biological Sciences
	Stanford University
	855 California Ave, Palo Alto, CA 94304, USA
	Tel: 650 723 2227
	Fax: 650 725 8221
	Email: walbot@stanford.edu
	Plate: 707014 row: F column: 11.
	Location/Qualifiers
	1..625
	/organism="Zea mays"
	/cultivar="W23"
	/db_xref="taxon:4577"
	/clone_lib="707 - Mixed adult tissues from Walbot lab (SK)"
	/tissue_type="tassel, kernel, silk, husk, root, leaf"
	/dev_stage="adult"
	/lab_host="DH10B"
	/note="Organ: tassel, kernel, silk, husk, root, leaf; Vector: pGAD10; Site_1: EcoRI; cDNA library from fully differentiated maize tissues from an active Mutator plant. Tissue ratio is 4/2/1/1/1 (tassel, kernel, silk, husk, root, leaf). Unidirectionally cloned."
BASE COUNT	131 a 181 c 175 g 138 t
ORIGIN	
Query Match	18.6%; Score 172.4; DB 72; Length 625;
Best Local Similarity	56.9%; Pred. No. 4.7e-42;
Matches	336; Conservative 0; Mismatches 251; Indels 3; Gaps 1;
7	ttaacctacatcttaccagacaatagctgtgtcccaacgltgtccacaacatcgttcgcgcacaca 66
28	TTGTTCCTCCACAGTCTCTATGAGCATTCTGAGGCCCAAGCGCAGAGATTTGTCAGTCCATT 87
67	atcgtcaacagagctacagatccgatccccaagatcgtgtcttcaatatattacgttgcacatc 126
88	GTGGCACAAGGCTGTGGCCNAGAGAGACAGAGATGGGGCATCTTTAGTCAGATGCATTTGC 147
127	catgactgtctctgtaatggtgtgagacgttagcatattacttgacaacacacacagatttc 186
148	CATGACTCTCTTTGTCAAGGCGTCCATGCTTTCGGTGTGTGGACAAACAGCAGCACATATA 207
187	cgcactgaaaagatgatcattcgcgagacgctlaacagcgccagagggcttltccagtatcgat 246
208	GTTACTGAGAAAGGTTCCAAACCCGAACAGAGACTCCCTCAGGGGGTTTGAAGTGAATCGAC 267
247	cgcatgaagagctgcgcgttgaatcagcatgacgatgccacaagatcagttgtgacagctgtctg 306
268	CAGATTAAAGGCTGCTTTGAGGCTCCTCCAGGACACAGTCTCTGTCGACACATTTGTT 327
307	actatagctgcgcacaacagagcgttgactctcttcgacgcgagacgcttcttgagagatccgc 366
328	GCCCTTTGGCGCTCGTGATTTCCACCGCCCTTGTTGGTGGACCATTACTGGAGAGCGCCACTT 387
367	ggtcagctgtacatcccaacaagcatctcttagatctctgccaacgaacgcaacttgctgtcca 426
388	GGCCGGAGAGACTCGTCGTGTCGAAGCATCCAGGGGCTCCAAACATGATCATCCAGCCCCC 447
427	tctctcaccctgtccccaagctgaagatatgactttagaacagttggtgtctgaatcgctcagat 486
448	AACCAACACTCTCCCACTATCATCAACCAAGTTTCAAGCGCCAGGGGCTCTCATGTGTG--T 504

QY	487	gacctgttgctctgtccggaggagacacacatlttgaaagaaccagtgtagatcatcatg	546
Db	505	GATGTTGTCGGCCCTCTCAGGTGGTGCACCACTTGTATGTCCTGGCTGCACAGTTTTCGG	564
QY	547	gatagcctacatcttcagcaacaccttggttacttgcacccacagctcga	596
Db	565	CAGAGGCTATACACACAGACAGCAGTATGGCATGGCTGCACACACACTGGA	614
RESULT	13		
LOCUS	AM278775		
DEFINITION	AM278775	691 bp mRNA EST	18-APR-2000
ACCESSION	AM278775	GI:6667324	
VERSION	AM278775.1		
KEYWORDS			
SOURCE	soybean.		
ORGANISM	Glycine max		
REFERENCE	Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eucosids I; Fabales; Fabaceae; Papilionoideae; Glycine.		
AUTHORS	Shoemaker, R., Kelm, P., Vodka, L., Expelding, J., Corryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harey, N., Schurr, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Meterson, R. and Wilson, R.		
TITLE	Public Soybean EST Project		
JOURNAL	Unpublished (1999)		
COMMENT	On Nov 22, 1999 this sequence version replaced g1:6461883. Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63144 for further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3224 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com Insert Length: 1137 Std Error: 0.00 Seq primer: -40RP from Gibco High quality sequence stop: 446. Location/Qualifiers 1. 691 /organism="Glycine max" /db_xref="taxon:3847" /clone="GENOME SYSTEMS CLONE ID: Gm-cl019-4060" /clone.lib="Gm-cl019" /issue.type="immature seed coats of greenhouse grown plants" /lab_host="DH10B (Gibco BRL)" /note="vector: pSPORT1 (Life Technologies); Site1: Not I; Site2: Sal I; This cDNA library was constructed from mRNA isolated from immature seed coats (200-300 mgs) of greenhouse grown plants. The library was prepared using the Life Technologies puperscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dt) sequence with a Not I restriction site. Sal I linker adapters were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells (Gibco BRL). This library was constructed by Dr. Lila Vodka and Dr. Anu Khanna."		
BASE COUNT	208 a	168 c	131 g
			182 t
			2 others





---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:35:18 ; Search time 76.18 Seconds

(Without alignments)  
96.075 Million cell updates/sec

Title: US-09-246-451-17

Perfect score: 1608

Sequence: 1 MQLPTPTFYDNCSPVNSIVR.....PLTGTCQIRLNCRFVNSNS 309

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A.Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1603	99.7	309	1 P95200	Horseradish peroxi
2	1598	99.4	308	1 P81167	Horseradish peroxi
3	1598	99.4	342	1 P95203	Horseradish peroxi
4	1589	98.8	308	1 P94810	Sequence of horser
5	1462	90.9	349	1 R32760	HRP/IGF-II fusion
6	921	57.3	352	1 W44931	Soybean seed coat
7	921	57.3	352	1 W73138	Soybean seed coat
8	768.5	47.8	333	1 W87893	Maize Pet's root pr
9	705.5	43.9	320	1 W38217	Stylosanthes humil
10	594.5	37.0	324	1 W16607	Soybean peroxidase
11	594.5	37.0	324	1 W81493	Soybean peroxidase
12	594.5	37.0	324	1 W67731	Soybean peroxidase
13	591.5	36.8	324	1 W16608	Soybean peroxidase
14	591.5	36.8	324	1 W81494	Soybean peroxidase
15	591.5	36.8	324	1 W67732	Soybean peroxidase
16	566	35.2	313	1 W16609	Soybean peroxidase
17	566	35.2	313	1 W81495	Soybean peroxidase
18	566	35.2	313	1 W67733	Soybean peroxidase
19	565	35.1	313	1 W16610	Soybean peroxidase
20	565	35.1	313	1 W81496	Soybean peroxidase
21	565	35.1	313	1 W67734	Soybean peroxidase
22	298.5	18.6	121	1 R32761	HRP/IGF-II fusion
23	113	7.0	20	1 R35215	Armorracia peroxida
24	111.5	6.9	1529	1 R41732	High molecular wei
25	111.5	6.9	1601	1 W30292	Non-Tyrosine Haemo
26	110	6.8	365	1 W12392	Collorus hirsutus m
27	107	6.7	20	1 R35213	Tobacco peroxidase
28	105	6.5	20	1 R35214	Brassica peroxidase
29	98	6.1	20	1 R35216	Potato peroxidase
30	95	5.9	372	1 R28843	Lignin peroxidase
31	94.5	5.8	317	1 W20966	H. pylori flagella
32	93.5	5.8	317	1 R23600	Recombinant hemato
33	92.5	5.8	456	1 W36050	Hybrid Marek's dis

34	91	5.7	1500	1 R30636	hCPSt. Carboxyl ph
35	90	5.6	560	1 W79954	Truncated adenovir
36	90	5.6	568	1 W79953	Truncated adenovir
37	90	5.6	573	1 W79952	Truncated adenovir
38	90	5.6	573	1 W79955	Truncated adenovir
39	90	5.6	581	1 W73419	Mutant adenovirus
40	90	5.6	581	1 W79950	Mutant adenovirus
41	90	5.6	581	1 W79951	Mutant adenovirus
42	90	5.6	581	1 W79949	Mutant adenovirus
43	90	5.6	581	1 W79905	Mutant adenovirus
44	90	5.6	581	1 W82732	Adenovirus type fl
45	89.5	5.6	336	1 R23077	Adenovirus wild-ty
					IL-3-G-CSF, recomb

## ALIGNMENTS

RESULT	1
ID	P95200
AC	P95200; standard; protein; 309 AA.
DE	18-JUL-1990 (first entry)
DE	Horseradish peroxidase gene (HRP) as encoded by synthetic gene
KW	Horseradish peroxidase; restriction sites; DNA construct;
KW	fusion protein; enzyme assays; immunoassays; synthetic gene.
OS	Horseradish.
FT	Key
FT	region
FT	location/Qualifiers
FT	1..161
FT	/note="encoded by 5' end half of synthetic HRP gene
FT	n90761"
FT	161..309
FT	/note="Coded for by 3' end half of synthetic HRP gene
FT	n94368"
FT	1..309
FT	/note="Coded for by synthetic HRP gene n90760"
FT	region
FT	W08903424-A.
FT	20-APR-1989.
FT	07-OCT-1987; G00833.
FT	08-OCT-1987; GH-023662.
FT	(RRBT-) Brit Bio-Tech Ltd.
FT	Edwards RM, Burke JF;
FT	WPI: 89-130042/17.
FT	N-PSDB: N90760, n90761; n94368.
FT	PT Synthetic gene encoding horseradish peroxidase -
FT	used for efficient prodn. of the enzyme or fusion prods. for
FT	use in enzyme assays, esp. immunoassays
FT	PS Disclosure: 45pp; English.
FT	CC The synthetic gene has many restriction sites. Also claimed is a DNA
FT	construct contg. bases 14-937. The construct can be a vector where the
FT	HRP gene and a desired gene are co-expressed as a single fusion prod.,
FT	a single polycistronic message, or 2 separate but linked transcriptional
FT	units. The fusion proteins are for use in enzyme assays, esp. immuno-
FT	assays. Genes carrying mutations that destroy HRP activity can be used
FT	to follow the frequency of reversion or suppression of the particular
FT	mutation introduced into the gene. The constructs can be used in
FT	transgenic animals, esp. mammals. The synthetic gene is prepd by
FT	coupling successive nucleotides and/or ligating appropriate oligomers.
FT	Sequence 309 AA:
FT	Query Match
FT	Best Local Similarity 99.7%; Score 1603; DB 1; Length 309;
FT	Matches 308; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
FT	QY 1 MQLPTPTFYDNCSPVNSIVRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDNTT 60
FT	
FT	1 MQLPTPTFYDNCSPVNSIVRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDNTT 60
FT	Db 1 MQLPTPTFYDNCSPVNSIVRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDNTT 60
FT	QY 61 SFRTEKDAFCNGANSARCPFYIDRKAAYESACPTTVCADLLTTAAQSYTLAGPSRWY 120
FT	
FT	61 SFRTEKDAFCNGANSARCPFYIDRKAAYESACPTTVCADLLTTAAQSYTLAGPSRWY 120
FT	Db 61 SFRTEKDAFCNGANSARCPFYIDRKAAYESACPTTVCADLLTTAAQSYTLAGPSRWY 120
FT	
FT	QY 121 PLGRDLSQAFDLANMNLPAFFTLPLQKDSFRNVGLNSSDVLVALSGGHTTEKNCGRF 180

```
Db 121 PIGRRSLQAFLDLANANLPAFFETLPQKDSFRNGLNSSDVLVALSGHTFGKNOCRF 180
181 IMRLYFNSMTGLPDPPLNTTYLQTLRGICPLNGNISALVDFDLRPTPTIDNKYYVLEE 240
Db 181 IMRLYFNSMTGLPDPPLNTTYLQTLRGICPLNGNISALVDFDLRPTPTIDNKYYVLEE 240
QY 241 QKGLIQSDQELFSSPDATDTIPLVRSFANSTQTFENAFVEMDMGNITPLTGTGQIRL 300
Db 241 QKGLIQSDQELFSSPDATDTIPLVRSFANSTQTFENAFVEMDMGNITPLTGTGQIRL 300
QY 301 NCRVNSNS 309
Db 301 NCRVNSNS 309

RESULT 2
P81167
ID P81167 standard; protein; 308 AA.
AC P81167;
DT 05-NOV-1990 (first entry)
DE Horseradish peroxidase.
KM horseradish peroxidase; Armoracia lapathifolia;
KW Armoracia rusticana; ss.
OS Armoracia sp.
PN J63207386-A.
PD 26-AUG-1988.
PF 19-FEB-1987; 037623.
PR 19-FEB-1987; JP-037623.
PA (TOYO) Toyobo KK.
DR N-PSDB; N81508.
PT New structural gene of horseradish peroxidase -
PT produced under artificially controlled conditions by gene
PT manipulation.
PS Disclosure: P: Japanese.
CC DNA was preferably isolated from tissue cultures of root cells of
CC Armoracia sp. The gene can be inserted into a plasmid for
CC transformation of suitable host cells. The product (HRP) can then
CC be collected from the culture medium.
CC See also N82195.
SQ Sequence 308 AA;

Query Match 99.4%; Score 1598; DB 1; Length 308;
Best Local Similarity 99.7%; Pred. No. 3.7e-155;
Matches 307; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QLTPTFYDNSCPVNSIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLDNNTS 61
Db 1 QLTPTFYDNSCPVNSIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLDNNTS 60
QY 62 FRTKEKDAFGANSAARGFPVIDRMAKAVERACPTVSCADLLTTAAOOSVTLAGPSMRVP 121
Db 61 FRTKEKDAFGANSAARGFPVIDRMAKAVERACPTVSCADLLTTAAOOSVTLAGPSMRVP 120
QY 122 LGRRDSIQAFLDLANANLPAFFETLPQKDSFRNGLNSSDVLVALSGHTFGKNOCRF 181
Db 121 LGRRDSIQAFLDLANANLPAFFETLPQKDSFRNGLNSSDVLVALSGHTFGKNOCRF 180
QY 182 MDRLYFNSMTGLPDPPLNTTYLQTLRGICPLNGNISALVDFDLRPTPTIDNKYYVLEE 241
Db 181 MDRLYFNSMTGLPDPPLNTTYLQTLRGICPLNGNISALVDFDLRPTPTIDNKYYVLEE 240
QY 242 KGLIQSDQELFSSPDATDTIPLVRSFANSTQTFENAFVEMDMGNITPLTGTGQIRL 301
Db 241 KGLIQSDQELFSSPDATDTIPLVRSFANSTQTFENAFVEMDMGNITPLTGTGQIRL 300
QY 302 CRVNSNS 309
Db 301 CRVNSNS 308
```

```
RESULT 3
ID P95203
AC P95203;
DT 18-JUL-1990 (first entry)
DE Horseradish peroxidase (HRP) as encoded by HRP gene modified for
DE expression in mammalian cells
DE Horseradish peroxidase; restriction sites; DNA construct;
KW fusion protein; enzyme assays; immunoassays; synthetic gene;
KW synthetic linker; pSD18.
OS Horseradish.
FH Key Location/Qualifiers
FT peptide 1..19
FT protein /note="N-terminal pre sequence"
FT peptide 20..327
FT peptide 328..342
FT peptide /note="C-terminal pro sequence"
PN MO8903424-A.
PD 20-APR-1989.
PF 07-OCT-1988; G00833.
PR 08-OCT-1987; GB-023662.
PA (BRI-) Brit Bio-Tech Ltd.
PI Edwards RM, Burke JF;
DR WPI; 89-130042/17.
DR N-PSDB; N90762.
PT Synthetic gene encoding horseradish peroxidase -
PT used for efficient produ. of the enzyme or fusion prods. for
PT use in enzyme assays, esp. immunoassays
PS Disclosure: 45pp; English.
CC The synthetic gene is prepd. by coupling successive nucleotides and/or
CC ligating appropriate oligomers. A construct containing the DNA which
CC encodes the mature protein (see FT) is claimed. The construct can be a
CC vector where the HRP gene and a desired gene are co-expressed as a single
CC fusion prod., a single polycistronic message, or 2 separate but linked
CC transcriptional units. The fusion proteins are for use in enzyme assays,
CC esp. immunoassays. Genes carrying mutations that destroy HRP activity
CC can be used to follow the frequency of reversion or suppression of
CC the particular mutation introduced into the gene. The constructs can
CC be used in transgenic animals, esp. mammals.
SQ Sequence 342 AA;
```

```
Query Match 99.4%; Score 1598; DB 1; Length 342;
Best Local Similarity 99.7%; Pred. No. 4.3e-155;
Matches 307; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QLTPTFYDNSCPVNSIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLDNNTS 61
Db 20 QLTPTFYDNSCPVNSIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLDNNTS 79
QY 62 FRTKEKDAFGANSAARGFPVIDRMAKAVERACPTVSCADLLTTAAOOSVTLAGPSMRVP 121
Db 80 FRTKEKDAFGANSAARGFPVIDRMAKAVERACPTVSCADLLTTAAOOSVTLAGPSMRVP 139
QY 122 LGRRDSIQAFLDLANANLPAFFETLPQKDSFRNGLNSSDVLVALSGHTFGKNOCRF 181
Db 140 LGRRDSIQAFLDLANANLPAFFETLPQKDSFRNGLNSSDVLVALSGHTFGKNOCRF 199
QY 182 MDRLYFNSMTGLPDPPLNTTYLQTLRGICPLNGNISALVDFDLRPTPTIDNKYYVLEE 241
Db 200 MDRLYFNSMTGLPDPPLNTTYLQTLRGICPLNGNISALVDFDLRPTPTIDNKYYVLEE 259
QY 242 KGLIQSDQELFSSPDATDTIPLVRSFANSTQTFENAFVEMDMGNITPLTGTGQIRL 301
Db 260 KGLIQSDQELFSSPDATDTIPLVRSFANSTQTFENAFVEMDMGNITPLTGTGQIRL 319
QY 302 CRVNSNS 309
Db 320 CRVNSNS 327

RESULT 4
P94810
```

ID P94810 standard; protein; 308 AA.  
AC P94810;  
DT 10-MAR-1993 (revised)  
DE 25-JUN-1989 (first entry)  
KW Sequence of horseradish peroxidase (HRP) Isoenzyme C; detection reagent;  
FH Superoxide radicals (SOD); ss.  
FT Key Location/Qualifiers  
FT cds 1..924  
PN EP-299682-A.  
PD 18-JAN-1989.  
PF 07-JUL-1988; 306222.  
PR 17-JUL-1987; GB-016938.  
PA (AMER-) Amersham Int. PLC.  
PI Chiswell DJ, Orlepp SA;  
DR WPI: 89-017184/03.  
DR N-PSDB: N93078.  
PT New synthetic DNA encoding horseradish peroxidase C Isoenzyme or variants  
PT or active fragments, useful eg as detection reagent.  
PS Disclosure; fig.6; 19pp; English.  
CC This sequence contains the horseradish peroxidase Isoenzyme C (HRP-C)  
CC complete sequence (309 AA's) and flanking AA's. Production is by recomb-  
CC inant techniques. A conservative variant can also be encoded by the  
CC corresponding DNA sequence by introducing base changes resulting in amino  
CC acid (AA) substitutions. Properties of this enzyme are essentially the  
CC same as the native HRP-C. A peptide contg. at least one HRP-C epitope can  
CC also be encoded by restriction enzyme cleavage of the same DNA sequence  
CC resulting in peptides of different lengths (contg. the active AA's Arg-183  
CC and Tyr-185). HRP-C is a scavenger of toxic superoxide radicals (SOD) and  
CC is useful as a detection reagent. It provides enhanced illumination by  
CC catalysis of a substrate such as luminol, in presence of a light-output  
CC enhancer.  
SQ Sequence 308 AA;

Query Match 98.8%; Score 1589; DB 1; Length 308;  
Best Local Similarity 99.4%; Pred. No. 3e-154;  
Matches 306; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 QLTPEFYNSCPNVSNIYRDTIVNELRSDPRIASILRLHFHDFVNGCASIILDDNTS 61  
DB 1 QLTPEFYNSCPNVSNIYRDTIVNELRSDPRIASILRLHFHDFVNGCASIILDDNTS 60  
QY 62 FRTEKDAFGNANSARGFPVIDRMKAAYESACPRIVSCADLLTTIAQOQSVTLAGGSPMRV 121  
DB 61 FRTEKDAFGNANSARGFPVIDRMKAAYESACPRIVSCADLLTTIAQOQSVTLAGGSPMRV 120  
QY 122 LGRDLSQAFLDLANANLPAEFFTLPOLKDSFRNVLNRSSDLVALSGHTEGKNOCRF 181  
DB 121 LGRDLSQAFLDLANANLPAEFFTLPOLKDSFRNVLNRSSDLVALSGHTEGKNOCRF 180  
QY 182 MDRLYNSNGLPDPTLNTTYLQTLRGCLPLNGMLSALYDFDLTPRTTFDNKYYVNEEQ 241  
DB 181 MDRLYNSNGLPDPTLNTTYLQTLRGCLPLNGMLSALYDFDLTPRTTFDNKYYVNEEQ 240  
QY 242 KGLIOSDDELFPSPDATDTIPLVRSFANSTOTFFNAFEADRMGNITPLATOGQIRLN 301  
DB 241 KGLIOSDDELFPSPDATDTIPLVRSFANSTOTFFNAFEADRMGNITPLATOGQIRLN 300  
QY 302 CRVNSNS 309  
DB 301 CRVNSNS 308  
RESULT 5  
ID R32760  
AC R32760;  
DT 23-JUN-1993 (first entry)  
DE HRP/IGF-II fusion protein.  
KW Insulin receptor; E.coli periplasmic protease resistance;  
KW wound healing; osteoporosis; muscle-wasting; growth disorder;  
KW horseradish peroxidase; inclusion bodies; ds.

OS Synthetic.  
FH Key Location/Qualifiers  
FH region 1..282  
FT /note="first 282 residues of HRP"  
FT protein 283..349  
FT /label= IGF-II  
PN M09303152-A.  
PD 18-FEB-1993.  
PF 27-JUL-1992; G01389.  
PR 29-JUL-1991; GB-016325.  
PR 05-FEB-1992; GB-002401.  
PA (BRBI-) BRITISH BIO-TECHNOLOGY LTD.  
PI Bowden L, Edwards RM;  
DR WPI: 93-076509/09.  
DR N-PSDB: 037614.  
PT New insulin-like Growth Factor II analogues - for treating growth  
PT disorders, osteoporosis and muscle wasting diseases, have less  
PT affinity for insulin receptors than unmodified IGF-II  
PS Example 4; Page 44-46; 71pp; English.  
CC This fusion protein comprises the first 282 amino acids of HRP  
CC fused to IGF-II via a methionine residue to allow for the release  
CC of mature IGF-II by cleavage with CNBr. The fusion lacks the final  
CC 27 amino acids of HRP. This fusion protein was not ideal for  
CC production of IGF-II; the IGF-II is only 19% of the total fusion  
CC protein and the HRP sequence contains two internal methionine  
CC residues which act as additional cleavage sites. The IGF-II is the  
CC wild-type sequence, i.e. with Arg residues at positions 37 and 38.  
CC The invention covers analogues of this sequence in which one or both  
CC of the Arg residues are subst. by other amino acids (pref. non-  
CC basic and esp. Gln). The mutants are protease-resistant and have  
CC reduced affinity for insulin receptors.  
SQ Sequence 349 AA;

Query Match 90.9%; Score 1462; DB 1; Length 349;  
Best Local Similarity 99.6%; Pred. No. 3.4e-141;  
Matches 281; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLPTEFYNSCPNVSNIYRDTIVNELRSDPRIASILRLHFHDFVNGCASIILDDNTT 60  
DB 1 MQLPTEFYNSCPNVSNIYRDTIVNELRSDPRIASILRLHFHDFVNGCASIILDDNTT 60  
QY 61 SFRTKDAFGNANSARGFPVIDRMKAAYESACPRIVSCADLLTTIAQOQSVTLAGGSPMRV 120  
DB 61 SFRTKDAFGNANSARGFPVIDRMKAAYESACPRIVSCADLLTTIAQOQSVTLAGGSPMRV 120  
QY 121 PLGRDLSQAFLDLANANLPAEFFTLPOLKDSFRNVLNRSSDLVALSGHTEGKNOCRF 180  
DB 121 PLGRDLSQAFLDLANANLPAEFFTLPOLKDSFRNVLNRSSDLVALSGHTEGKNOCRF 180  
QY 181 IMDRLYNSNGLPDPTLNTTYLQTLRGCLPLNGMLSALYDFDLTPRTTFDNKYYVNEEQ 240  
DB 181 IMDRLYNSNGLPDPTLNTTYLQTLRGCLPLNGMLSALYDFDLTPRTTFDNKYYVNEEQ 240  
QY 241 OKGLIOSDDELFPSPDATDTIPLVRSFANSTOTFFNAFEAM 282  
DB 241 OKGLIOSDDELFPSPDATDTIPLVRSFANSTOTFFNAFEAM 282  
RESULT 6  
ID W44931  
AC W44931;  
DT 09-NOV-1998 (first entry)  
DE Soybean seed coat peroxidase.  
KW Peroxidase; seed coat; soybean; transgenic plant.  
OS Glycine max cv. Harosoy 63.  
FH Key Location/Qualifiers  
FH Peptide 1..26  
FH /label= Sig\_peptide  
FH Protein 29..352  
FH /label= Mat\_protein  
FT Active\_site 59..70

FT Domain 188..197  
/Label= Haem-binding  
CA2211018-A.  
PN 30-MAR-1998.  
PD 19-SEP-1997; 211018.  
PR 30-SEP-1996; US-722414.  
PR 30-SEP-1996; CA-186833.  
PA (MIAC ) CANADA MIN AGRICULTURE.  
PI Gijzen M.  
DR WPI: 98-388720/34.  
DR N-PSDB: V30460-61.  
PT New DNA used in preparation of soybean seed coat peroxidase - used  
PT e.g. in diagnostic assays, for de-linking waste paper, for oxidation  
PT of primary alcohols, in paper and pulp industries and to encode  
PT herbicide resistance in plants  
PS Example 1: Page 42-45: 75pp: English.  
CC This is the amino acid sequence of soybean seed coat peroxidase.  
CC Its was deduced from isolated cDNA and genomic DNA clones (see  
CC V30460-61). Vectors, host cells, and transgenic plants are  
CC provided, as well as methods for the production of recombinant  
CC seed coat peroxidase. Soybean coat peroxidases are useful  
CC industrially, e.g. their high thermal stability even at extremes of  
CC pH make them useful in diagnostic assays as an enzyme label for  
CC e.g. antigens, and within staining techniques. They are also  
CC useful for de-linking printed waste paper, biocatalytic oxidation  
CC of primary alcohols, as a replacement for chlorine in the pulp and  
CC paper industry or as a formaldehyde replacement for use in  
CC adhesives.  
SQ Sequence 352 AA:

Query Match 57.3%; Score 921; DB 1; Length 352;  
Best Local Similarity 59.1%; Pred. No. 5.8e-86;  
Matches 182; Conservative 40; Mismatches 84; Indels 2; Gaps 2;

QY 2 QLPPTFDNSCPNVSNIVRDTIVNELRSDPRIASILRLHFHFCFVNGCDASILLDNNTTS 61  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
Db 27 QLPPTFYRETCPNLFPVFGVIFDASFTDPRIGASLMRLHFHDFVOGCGSVLNNNTDF 86  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
QY 62 FRTKEKAFGANANRGPVVDKRAAVESACPTVSCADLLTTAAQOSVTLAGPSPMRVP 121  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
Db 87 IESQDLPININSIRGLDVNDIKTAVENSCPTVSCADLLTAAEIASVYLGSGPMPVP 146  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
QY 122 LGRDLSQAFLDLANANLPAPFETLPQKDSFRNNGNRSDDLVALSGHTFGKNCRFI 181  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
Db 147 LGRDLSLTANRTLANONLPAPFETLPQKDSFAVOGLN-TLIDLVTLSGHTFGRCSTF 205  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
QY 182 MDRLYNSNTGLPDPPLNTTYLQTLRGCLPGLNGLSALVPDFLRTPTTFDNKYYVNLREQ 241  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
Db 206 INRLYNFSNTGNPDPTLNTTYLLEVLARCPQMATGDMLTNLDLSTPDQFNNRYSNLQL 265  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
QY 242 KGLISQDELFSSPDATDTPLVRSFANSQTGFENAFVEMADRMGNTTPTLTGQGIIRLN 301  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
Db 266 NGLISQDELFSTPGA-DTILPIVNSFSSNONTFFSNFRVSMIKMGNTIGVLTGDEGEIRLQ 324  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
QY 302 CRVYNSNS 309  
| | | : | | : |  
Db 325 CNFVNGDS 332

RESULT 7  
W73138  
ID W73138 standard; Protein: 352 AA.  
AC W73138;  
DT 02-FEB-1999 (first entry)  
DE Soybean seed coat peroxidase.  
KW Peroxidase; seed coat; soybean; transgenic plant.  
OS Glycine max cv. Harosoy 63.  
FH Key Location/Qualifiers  
FT 1..26  
FT /label= Sig\_peptide  
FT 27..352  
FT Protein /label= Mat\_protein

FT Active\_site 59..70  
FT Domain 188..197  
/note= "haem-binding domain"  
CA2186833-A.  
PN 31-MAR-1998.  
PD 30-SEP-1996; 186833.  
PR 30-SEP-1996; CA-186833.  
PA (MIAC ) CANADA MIN AGRICULTURE.  
PI Gijzen M.  
DR WPI: 98-521791/45.  
DR N-PSDB: V59077-78.  
PT New DNA encoding soybean seed-coat peroxidase and the promoter - useful  
PT for expression of heterologous genes in the seed coat, which may  
PT give herbicide resistance, viral protection and pest and pathogen  
PT control, and may improve the taste or nutritional value of soybean  
PT beans.  
PS Example 1: Fig 1; 61pp: English.  
CC This is the amino acid sequence of a soybean seed coat peroxidase,  
CC as deduced from an isolated cDNA clone sequence (see V59077).  
CC Soybean plants containing a dominant bp allele accumulate large  
CC amounts of peroxidase in the hourglass cells of the subepidermis.  
CC Homozygous recessive eep genotypes are much reduced in total seed  
CC coat peroxidase activity. Soybean peroxidases have been shown to  
CC have useful catalytic properties and a high degree of thermal  
CC stability even at extremes of pH. They are used as enzyme labels  
CC in diagnostic assays. The invention relates to soybean seed coat  
CC polynucleotides, and to regulatory regions of the peroxidase gene  
CC that can be used to control the expression of genes of interest  
SQ Sequence 352 AA:

Query Match 57.3%; Score 921; DB 1; Length 352;  
Best Local Similarity 59.1%; Pred. No. 5.8e-86;  
Matches 182; Conservative 40; Mismatches 84; Indels 2; Gaps 2;

QY 2 QLPPTFDNSCPNVSNIVRDTIVNELRSDPRIASILRLHFHFCFVNGCDASILLDNNTTS 61  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
Db 27 QLPPTFYRETCPNLFPVFGVIFDASFTDPRIGASLMRLHFHDFVOGCGSVLNNNTDF 86  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
QY 62 FRTKEKAFGANANRGPVVDKRAAVESACPTVSCADLLTTAAQOSVTLAGPSPMRVP 121  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
Db 87 IESQDLPININSIRGLDVNDIKTAVENSCPTVSCADLLTAAEIASVYLGSGPMPVP 146  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
QY 122 LGRDLSQAFLDLANANLPAPFETLPQKDSFRNNGNRSDDLVALSGHTFGKNCRFI 181  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
Db 147 LGRDLSLTANRTLANONLPAPFETLPQKDSFAVOGLN-TLIDLVTLSGHTFGRCSTF 205  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
QY 182 MDRLYNSNTGLPDPPLNTTYLQTLRGCLPGLNGLSALVPDFLRTPTTFDNKYYVNLREQ 241  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
Db 206 INRLYNFSNTGNPDPTLNTTYLLEVLARCPQMATGDMLTNLDLSTPDQFNNRYSNLQL 265  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
QY 242 KGLISQDELFSSPDATDTPLVRSFANSQTGFENAFVEMADRMGNTTPTLTGQGIIRLN 301  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
Db 266 NGLISQDELFSTPGA-DTILPIVNSFSSNONTFFSNFRVSMIKMGNTIGVLTGDEGEIRLQ 324  
| | | | | : | | | : | | | : | | | : | | | : | | | : | | | :  
QY 302 CRVYNSNS 309  
| | | : | | : |  
Db 325 CNFVNGDS 332

RESULT 8  
W87893  
ID W87893 standard; Protein: 333 AA.  
AC W87893;  
DT 12-APR-1999 (first entry)  
DE Maize Per5 root preferential cationic peroxidase.  
KW Peroxidase; Per5; maize; corn; transgenic plant; promoter;  
KW root.  
OS Zea mays.  
FH Key Location/Qualifiers  
FT 1..32  
FT Peptide /note= "signal peptide"

FT Protein 33.333  
FT /note= "mature protein"  
FT Modified\_site 85  
FT /note= "putative N-glycosylation"  
FT Modified\_site 170  
FT /note= "putative N-glycosylation"  
FT Modified\_site 213  
FT /note= "putative N-glycosylation"  
FT Modified\_site 311  
FT /note= "putative N-glycosylation"  
PN W09856921-A1.  
PD 17-DEC-1998.  
PF 10-JUN-1998; U11921.  
PR 12-JUN-1997; US-049752.  
PA (DOMC ) DOM AGROSCIENCES LLC.  
PI Alanley M, Armstrong K, Belmar S, Folkerts O, Hopkins N,  
PI Menke MA, Pareddy D, Petolino JF, Smith K, Woosley A.  
DR WPI: 99-080904/07.  
DR N-PSDB: V63717 AND V63720.  
PT New isolated regulatory sequences for transgenic plants - which are  
PT derived from the maize root preferential cationic peroxidase protein  
PT (per5) gene  
PS Disclosure: Page 89-90. 150pp: English.  
CC This is the amino acid sequence of peroxidase per5 of maize. The  
CC enzyme is encoded by the root preferential cationic peroxidase  
CC per5 gene (see V63717). Regulatory sequences of the per5 gene,  
CC including the promoter, introns and 3' untranslated region (3'UTR),  
CC are used in claimed recombinant gene cassettes for controlling  
CC expression of recombinant genes in selected tissue, especially  
CC the root, of transformed plants, particularly maize. The gene  
CC cassettes can be used for expression of heterologous genes such as  
CC those that confer tolerance to herbicides, insects or viruses, and  
CC characteristcs to provide improved nutritional value or processing  
SQ Sequence 333 AA:

Query Match 47.8%; Score 768.5; DB 1; Length 333;

Best Local Similarity 49.7%; Pred. No. 2e-70;

Matches 152; Conservative 52; Mismatches 99; Indels 3; Gaps 2;

QY 3 LTPFTYDSCPNVSNIVRTIVNELRSDPRIASILRLHFHDCVNGCDASILIDNTTSF 62  
DB 30 LEPFYHSCCKAREIYOSIVAOAVAKETRMASILVRLHFHDCVNGCDASVILIDNTSSSI 89  
QY 63 RTEKDAFGNANSARGPEYIDRMKAAVESACPRVSCADLITIAQOSVTLAGPSPWRVP 122  
DB 90 VSEKGSNPNRNSLGFEEYIDQIKALEACPGTVSCADIVALARSTVALGCGPYMDVPL 149  
QY 123 GRBDSLOAFDLANANLPAPFPTLPOLKDSFRNVLGRSSDVALSGHTEGKNOCRFIM 182  
DB 150 GRBDSLAGSIGSNNDIFAPNNTLPTITTRKROGLN-VVDVVALSGHTEGMSCTISFR 208  
QY 183 DRLYNFSNTGLPDPPLNTLTYLQTLRGCLPLNGNL-SALVDPDLRPTTIFDNKYVYVLEBQK 242  
DB 209 QRLYNQNGNAGDSTLDPLVYAAKLROCGPRSGDNMLPLDLITPAKDNFYKKLNLG 268  
QY 243 GLIYSDQELFSSPDATDITPLVRSFANSTOTFFNAFYAAMDRMGNTITPLTGOIGRLNC 302  
DB 269 GLIYSDDELTLRK--SAETAAALVAKYAADVNLFOHFASQVNMGNISPLTSGSGEIRKNC 326  
QY 303 RVVNSN 308  
DB 327 RRLNND 332

RESULT 9

W38217 M38217 standard: Protein; 320 AA.

AC W38217;

DT 27-APR-1998 (first entry)

DE Stylosanthes humilis Shpx6 peroxidase.

KW Peroxidase; Shpx6; transgenic plant; fungus resistance;

KW disease protection; Phytophthora parasitica;  
KW Leptosphaeria maculans; Sclerotinia sclerotiorum.  
OS Stylosanthes humilis strain Paterson.  
FH Key Location/Qualifiers  
FT Peptide 1..24  
FT /label= Sig-peptide  
PN W09741237-A1.  
PD 06-NOV-1997.  
PF 29-APR-1997; AU0253.  
PR 29-APR-1996; AU-009532.  
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.  
PI Goulter KC, Kazan K, Manners JM.  
DR WPI: 97-549739/50.  
DR N-PSDB: T95782.  
PT Plant cell transformed with Stylosanthes humilis Shpx6 peroxidase -  
PT useful to confer on plants resistance to fungi, e.g. Phytophthora  
PT parasitica, Leptosphaeria maculans or Sclerotinia sclerotiorum  
PS Claim 4; Page 15-16; 30pp: English.  
CC This protein comprises the Shpx6 peroxidase of Stylosanthes  
CC humilis. A claimed method of engineering a plant to fungal  
CC resistance comprises introducing into cells of the plant a DNA  
CC construct comprising a promoter that is constitutively operative  
CC in the plant cell (preferably the cauliflower mosaic virus 35S  
CC promoter) and a Shpx6 DNA sequence (see T95782), or a hybridising  
CC sequence or fragment that encodes an enzyme with peroxidase  
CC activity. Also claimed are plant cells harboring the DNA  
CC construct, a plant comprising such cells, and material from such  
CC a plant, especially seed, pollen, a stem segment or a cutting.  
CC Resistance to fungi, especially Phytophthora parasitica,  
CC Leptosphaeria maculans or Sclerotinia sclerotiorum, can be  
CC conferred on cereal, legume, oilseed, sugar or fibre plants,  
CC particularly maize, banana, peanut, field pea, sunflower, tomato,  
CC carola, tobacco, wheat, barley, oat, potato, soybean, cotton,  
CC carnation and sorghum.  
SQ Sequence 320 AA:

Query Match 43.9%; Score 705.5; DB 1; Length 320;

Best Local Similarity 48.4%; Pred. No. 5e-64;

Matches 149; Conservative 51; Mismatches 93; Indels 15; Gaps 6;

QY 2 QLTPTFYDSCPNVSNIVRTIVNELRSDPRIASILRLHFHDCVNGCDASVILIDNTTS 61  
DB 25 QLSNFEVATTCPNMLSTRSGVNSAVSKARMGASLILRLHFHDCVNGCDASVILIDNTSN 84  
QY 62 RTEKDAFGNANSARGPEYIDRMKAAVESACPRVSCADLITIAQOSVTLAGPSPWRVP 121  
DB 85 FTGEXTARPNANSIRGEFVIDTIKRSQVESLCPGVYSCADILAVARSDVVALGPGSWTVQ 144  
QY 122 LGRBDSLOAFDLANANLPAPFPTLPOLKDSFRNVLGRSSDVALSGHTEGKNOCRFI 181  
DB 145 LGRDSTTASJSLANSDDLAAPTLDLISLISAFSKKGLS-TSEMVALLSGHTEGQARCTSF 203  
QY 182 MDRLYNFSNTGLPDPPLNTLTYLQTLRGCLPL---NGNL-SALVDPDLRPTTIFDNKYVYVNL 238  
DB 204 RTRITTESN-----IDPNFAKSLQGNCPMTGTGND-NMLAPIDTISPTFEDNKTYVNL 255  
QY 239 EEQKGLIYSDQELFSSPDATDITPLVRSFANSTOTFFNAFYAAMDRMGNTITPLTGOIGOI 298  
DB 256 LVYKGLFHSDDQLFNG-GSTDS--QVNGYASNPSSFCDFENAMIKMGNISPLTSGSGOI 312  
QY 299 RLNCRVYN 306  
DB 313 RTNCRKTN 320

RESULT 10

W16607 M16607 standard: Protein; 324 AA.

AC W16607;

DT 25-JUL-1997 (first entry)

DE Soybean peroxidase Sepal.



KM Peroxidase; SEpal; soybean; transgenic plant; immunoassay; assay;  
KW marker; pulp; paper; bleaching; bioremediation;  
OS soil decontamination.  
FH Key Location/Qualifiers  
FT Peptide 1..21 /note= "signal peptide"  
FT Protein 22..324 /note= "mature protein"  
FT Peptide 22..28 /note= "QLXXFY motif, characteristic of plant peroxidases"  
FT Disulfide\_bond 32 /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 65 /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 70 /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 110 /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 116 /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 195 /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 227 /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 319 /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 319 /note= "Cys residue involved in disulfide bonding"  
FT Modified\_site 77 /note= "N-glycosylation"  
FT Modified\_site 90 /note= "N-glycosylation"  
FT Modified\_site 149 /note= "N-glycosylation"  
FT Modified\_site 163 /note= "N-glycosylation"  
FT Modified\_site 204 /note= "N-glycosylation"  
FT Modified\_site 235 /note= "N-glycosylation"  
FT Modified\_site /note= "N-glycosylation"  
PN W09855629-A2.  
PD 10-DIC-1998.  
PE 04-JUN-1998; U11421.  
PR 04-JUN-1997; US-868577.  
PA (INDI-) INDIANA CROP IMPROVEMENT ASSOC.  
PI Vierling RA:  
DR WPI; 99-070273/06.  
DR N-PSDB; V81418 AND V81420.  
PT New soybean peroxidase genes - useful, e.g. in pulp and paper bleaching, on site waste destruction and soil remediation  
PS Example 8; Page 38-40; 78pp; English.  
CC This is the amino acid sequence of novel soybean peroxidase SEpal, as deduced from a cDNA clone (see V81420) isolated from a soybean cv. Resnick seedling cDNA library. Soybean peroxidase SEpal, SEpa2, SEpb1 and SEpb2 (see W67731-34) sequences are provided.  
CC SEpal and SEpa2 exhibit 97% amino acid identity, SEpb1 and SEpb2 share 95% amino acid identity, and SEpal and SEpb1 share 47% amino acid identity. The peroxidases are useful in pulp and paper bleaching, on-site waste destruction, soil remediation, organic synthesis and diagnostic chemistries. Soybean peroxidase has advantages over chlorine bleach, being cheaper, more environmentally friendly, and producing hydroxyl ions with twice the oxidising power of chlorine ions. The plant enzyme is cheap and easy to produce. The invention also relates to immunoassays or oligonucleotide assays which utilise soybean peroxidase as marker.  
CC Sequence 324 AA:  
QY Query Match 37.0%; Score 594.5; DB 1; Length 324;  
Best Local Similarity 39.9%; Pred. No. 1,1e-52;  
Matches 122; Conservative 63; Mismatches 115; Indels 7; Gaps 5;

Db 22 QLOGFYANSCPKAEQIVLKFEVHDIHNPISLAALIFMHPHDCFVRCDASVILNSTN 81  
QY 62 FRTEKDAFGNANSARGFPEVIDRMKAIVESACPRVSCADLTITAAOOSTYLAGPSMPVP 121  
Db 82 -QAEKNAPPNL-TVKGPEFDITIKSLVEACPGVASCADLTITLARDIVAVGPGMPVP 139  
QY 122 LGRDSTLQAFDILANANLPAPFETLPOLKDSFRVNGVLRSSDVALSGCHFGKNOCFI 181  
Db 140 TGRDGVANSNTEARNNIPAPSSNFTTQTLEFANQGLD-LKDVLLSNAHTIGAHCSL 198  
QY 182 MDRLYNFSNGLPPELTNTTYLQTLRGL-CPLNGNSALVDPLRPTTFDMKRYVNL 239  
Db 199 SNRLFNFTGKGDDPSLDSEVAANLKAFCYDLKMLNTTKIEMPGSKTFEDLSYSHVI 258  
QY 240 EQKGLIQSDQELFSPQATDITIPVRSANSTQFFNAFVAMBRMGITPLTGOQIR 299  
Db 259 KRRLFESDAALTR-NSVTKAOITILLEGSENFFAFATSIKMGRIHVKTGEGR 316  
QY 300 LNCRVNS 307  
Db 317 KHCAPINS 324  
RESULT 13  
W16608  
ID W16608 standard; Protein; 324 AA.  
AC W16608:  
DT 25-JUL-1997 (first entry)  
DE Soybean peroxidase SEpa2.  
KW waste disposal; SEpa2; soybean; plant breeding; pulp; paper; bleaching;  
OS Glycine max cv. Resnik.  
FH Key Location/Qualifiers  
FT Peptide 1..21 /label= Sig\_peptide  
FT Protein 22..324 /label= Mat\_protein  
FT Peptide 22..28 /label= QLXXFY\_motif  
FT binding\_site 61..67 /label= Haem\_binding\_site  
FT modified\_site 77 /label= N-glycosylation\_site  
FT modified\_site 90 /label= N-glycosylation\_site  
FT modified\_site 149 /label= N-glycosylation\_site  
FT modified\_site 163 /label= N-glycosylation\_site  
FT modified\_site 204 /label= N-glycosylation\_site  
FT modified\_site 235 /label= N-glycosylation\_site  
FT modified\_site /label= N-glycosylation\_site  
PN W09715656-A1.  
PD 01-MAY-1997.  
PE 11-OCT-1996; U16354.  
PR 27-OCT-1995; US-549658.  
PA (INDI-) INDIANA CROP IMPROVEMENT ASSOC.  
PI Vierling RA:  
DR WPI; 97-259011/23.  
DR N-PSDB; T66685.  
PT Isolated soybean peroxidase genes - useful for developing products for quantification and monitoring of peroxidase activity  
PS Claim 5; Page 36-37; 63pp; English.  
CC Soybean peroxidases SEpal, SEpa2, SEpb1 and SEpb2 (W16607-10) are respectively encoded by cDNA clones (T66684-87) obtd. from a soybean seed bud cDNA library. Isolation of these clones allows the prodn. of the soybean peroxidases, free of other soybean proteins, in transformed host cells. The peroxidases can be used to raise monoclonal antibodies (MAb). A MAb, which is preferably immunoreactive with the peroxidase and does not interfere with its enzymatic activity when bound, can be used in a non-destructive





PR 04-JUN-1997; US-868577.  
PA (INDI-) INDIANA CROP IMPROVEMENT ASSOC.  
PI Wierling RA;  
PI Wierling RA;  
DR MPI; 99-070273/06.  
DR N-PSDB; V81421.  
PT New soybean peroxidase genes - useful, e.g. in pulp and paper  
bleaching, on site waste destruction and soil remediation  
Example 8; Page 41-43; 78pp: English.  
CC This is the amino acid sequence of novel soybean peroxidase SEP2,  
as deduced from a cDNA clone (see V81421) isolated from a soybean  
cv. Resnick seedling cDNA library. Soybean peroxidase SEP1,  
SEP2, SEP1 and SEP2 (see W67731-34) sequences are provided.  
CC SEP1 and SEP2 exhibit 97% amino acid identity, SEP1 and SEP2  
share 95% amino acid identity, and SEP1 and SEP1 share 47% amino  
acid identity. The peroxidases are useful in pulp and paper  
bleaching, on-site waste destruction, soil remediation, organic  
synthesis and diagnostic chemistries. Soybean peroxidase has  
advantages over chlorine bleach, being cheaper, more environmentally  
friendly, and producing hydroxyl ions with twice the oxidising  
power of chlorine ions. The plant enzyme is cheap and easy to  
produce. The invention also relates to immunoassays or  
oligonucleotide assays which utilise soybean peroxidase as marker.  
CC Sequence 324 Aa.

Query Match	36.8%;	Score 591.5;	DB 1;	Length 324;
Best Local Similarity	39.9%;	Pred. No. 2.2e-52;		
Matches 123;	Conservative 62;	Mismatches 116;	Indels 7;	Gaps 5

[illegible]

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:38:09 ; Search time 66.92 Seconds  
(without alignments)  
77.394 Million cell updates/sec

Title: US-09-246-451-17  
Perfect score: 1608  
Sequence: 1 MOLPTFPYDNCSPVSNIVR.....PLTGTQGIRLNCRVNSNS 309

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 1676186 residues  
Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgnl\_7/ptodata/1/1aa/5A-COMB.pep:\*  
2: /cgnl\_7/ptodata/1/1aa/5B-COMB.pep:\*  
3: /cgnl\_7/ptodata/1/1aa/6-COMB.pep:\*  
4: /cgnl\_7/ptodata/1/1aa/PCTUS-COMB.pep:\*  
5: /cgnl\_7/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1462	90.9	349	US-08-190-029A-10	Sequence 10, Appl
2	1462	90.9	349	US-08-462-695-10	Sequence 10, Appl
3	594.5	37.0	324	US-08-671-320-11	Sequence 11, Appl
4	594.5	37.0	324	US-08-868-577-11	Sequence 11, Appl
5	591.5	36.8	324	US-08-671-320-13	Sequence 13, Appl
6	591.5	36.8	324	US-08-868-577-13	Sequence 13, Appl
7	566	35.2	313	US-08-671-320-15	Sequence 15, Appl
8	566	35.2	313	US-08-868-577-15	Sequence 15, Appl
9	565	35.1	313	US-08-671-320-17	Sequence 17, Appl
10	565	35.1	313	US-08-868-577-17	Sequence 17, Appl
11	298.5	18.6	121	US-08-190-029A-12	Sequence 12, Appl
12	298.5	18.6	121	US-08-462-695-12	Sequence 12, Appl
13	118.5	7.4	726	US-08-313-185-49	Sequence 49, Appl
14	118.5	7.4	726	US-08-459-499-13	Sequence 13, Appl
15	116	7.2	1600	US-08-617-697-10	Sequence 10, Appl
16	111	6.9	1529	US-08-728-470-10	Sequence 10, Appl
17	110.5	6.9	294	US-08-459-499-16	Sequence 16, Appl
18	110.5	6.9	729	US-08-313-185-50	Sequence 50, Appl
19	110.5	6.9	729	US-08-459-499-14	Sequence 14, Appl
20	107	6.7	293	US-08-313-185-52	Sequence 52, Appl
21	85.5	5.3	1358	US-08-404-665-4	Sequence 4, Appl
22	85.5	5.3	1358	US-08-404-671-4	Sequence 4, Appl
23	85.5	5.3	1358	US-08-404-781-4	Sequence 4, Appl
24	84	5.2	560	US-08-756-317-8	Sequence 8, Appl
25	84	5.2	933	US-08-293-728-2	Sequence 2, Appl
26	83.5	5.2	928	US-08-474-140-11	Sequence 11, Appl
27	83.5	5.2	928	US-08-477-630-11	Sequence 11, Appl
28	83.5	5.2	928	US-08-472-293-11	Sequence 11, Appl

29	83.5	5.2	928	US-08-474-545-11	Sequence 11, Appl
30	83.5	5.2	928	US-08-478-341-11	Sequence 11, Appl
31	83.5	5.2	928	US-08-996-733-11	Sequence 11, Appl
32	82	5.1	553	US-08-663-566A-13	Sequence 13, Appl
33	82	5.1	553	US-08-484-575A-14	Sequence 14, Appl
34	82	5.1	553	US-08-023-610-13	Sequence 13, Appl
35	82	5.1	553	US-08-288-065A-13	Sequence 13, Appl
36	82	5.1	553	US-08-362-240A-13	Sequence 13, Appl
37	82	5.1	553	US-08-477-459-14	Sequence 14, Appl
38	82	5.1	553	PCT-US94-01826A-14	Sequence 14, Appl
39	82	5.1	553	PCT-US94-02252A-14	Sequence 14, Appl
40	82	5.1	553	PCT-US95-10245-13	Sequence 13, Appl
41	81	5.0	638	US-08-426-125-5	Sequence 5, Appl
42	81	5.0	638	US-08-455-355-5	Sequence 5, Appl
43	80	5.0	313	US-08-373-858-2	Sequence 2, Appl
44	80	5.0	313	US-08-500-611-2	Sequence 2, Appl
45	80	5.0	313	US-08-500-694-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-190-029A-10  
Sequence 10, Application US/08190029A  
Patent No. 5736363  
GENERAL INFORMATION:  
APPLICANT: EDWARDS, Richard Mark  
APPLICANT: BAWDEN, Lindsey  
TITLE OF INVENTION: IGF-II ANALOGUES  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: ALLEGRETTI & MITCHELL, LTD.  
STREET: 10 S. WACKER DRIVE, SUITE 3000  
CITY: CHICAGO  
STATE: ILLINOIS  
COUNTRY: U.S.A.  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/190,029A  
FILING DATE: 28-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/G892/01389  
FILING DATE: 27-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9202401.7  
FILING DATE: 05-FEB-1992  
APPLICATION NUMBER: GB 9116325.3  
FILING DATE: 29-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: JOHN J. McDONNELL  
REGISTRATION NUMBER: 26,949  
REFERENCE/DOCKET NUMBER: 94,062  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 349 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-190-029A-10  
Query Match 90.9%; Score 1462; DB 1; Length 349;

Best Local Similarity 99.6%; Pred. No. 3e-152;  
Matches 281; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 MQLTFEYDNCSPNVSIVADTVINELRSDPRIAASILRLHFHDCFFVNGCDASILLDNTT 60
Db 1 MQLTFEYDNCSPNVSIVADTVINELRSDPRIAASILRLHFHDCFFVNGCDASILLDNTT 60
OY 61 SFRTKDAFGNANSARGFPVIDRMKAAVESACPRVSCADLLTTAAQSVTLAGGPSMRV 120
Db 61 SFRTKDAFGNANSARGFPVIDRMKAAVESACPRVSCADLLTTAAQSVTLAGGPSMRV 120
OY 121 PLGRRSLOAFDLAANLPAPEFTLPQLKDSFRNGLNSSLVALSGGHTGKNCRF 180
Db 121 PLGRRSLOAFDLAANLPAPEFTLPQLKDSFRNGLNSSLVALSGGHTGKNCRF 180
OY 181 IMRLVFSNTGLPDPPLNTTYLQTLRGICPLNGNLSALVDFDLRPTTFIDNKYYVLEE 240
Db 181 IMRLVFSNTGLPDPPLNTTYLQTLRGICPLNGNLSALVDFDLRPTTFIDNKYYVLEE 240
OY 241 QKGLIQSDQELFSSPDATDTIPLVRSFANSTQTFENAFVEAM 282
Db 241 QKGLIQSDQELFSSPDATDTIPLVRSFANSTQTFENAFVEAM 282
```

```
RESULT 2
US-08-462-695-10
; Sequence 10, Application US/08462695
; Patent No. 5854025
; GENERAL INFORMATION:
; APPLICANT: EDWARDS, Richard Mark
; APPLICANT: BANDEN, Lindsey
; TITLE OF INVENTION: IGF-II ANALOGUES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER 6 ALLEGRETTI, LTD.
; STREET: 10 S. WACKER DRIVE, SUITE 3000
; CITY: CHICAGO
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,695
; FILING DATE: 5-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/190,029
; FILING DATE: 28-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB92/01389
; FILING DATE: 27-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9202401.7
; FILING DATE: 05-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9116325.3
; FILING DATE: 29-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: JOHN J. McDONNELL
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 94,062-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1234
; TELEFAX: 312-715-1000
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 349 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

; MOLECULE TYPE: protein  
US-08-462-695-10

Query Match 90.9%; Score 1462; DB 2: Length 349;  
Best Local Similarity 99.6%; Pred. No. 3e-152;  
Matches 281; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 MQLTFEYDNCSPNVSIVADTVINELRSDPRIAASILRLHFHDCFFVNGCDASILLDNTT 60
Db 1 MQLTFEYDNCSPNVSIVADTVINELRSDPRIAASILRLHFHDCFFVNGCDASILLDNTT 60
OY 61 SFRTKDAFGNANSARGFPVIDRMKAAVESACPRVSCADLLTTAAQSVTLAGGPSMRV 120
Db 61 SFRTKDAFGNANSARGFPVIDRMKAAVESACPRVSCADLLTTAAQSVTLAGGPSMRV 120
OY 121 PLGRRSLOAFDLAANLPAPEFTLPQLKDSFRNGLNSSLVALSGGHTGKNCRF 180
Db 121 PLGRRSLOAFDLAANLPAPEFTLPQLKDSFRNGLNSSLVALSGGHTGKNCRF 180
OY 181 IMRLVFSNTGLPDPPLNTTYLQTLRGICPLNGNLSALVDFDLRPTTFIDNKYYVLEE 240
Db 181 IMRLVFSNTGLPDPPLNTTYLQTLRGICPLNGNLSALVDFDLRPTTFIDNKYYVLEE 240
OY 241 QKGLIQSDQELFSSPDATDTIPLVRSFANSTQTFENAFVEAM 282
Db 241 QKGLIQSDQELFSSPDATDTIPLVRSFANSTQTFENAFVEAM 282
```

```
RESULT 3
US-08-671-320-11
; Sequence 11, Application US/08671320
; Patent No. 5840558
; GENERAL INFORMATION:
; APPLICANT: VIERLING JR, RICHARD A
; TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN
; ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 1125 SO. 103RD STREET
; STREET: SUITE 330
; CITY: OMAHA
; STATE: NE
; COUNTRY: US
; ZIP: 68124-1076
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/671,320
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JONDLE, ROBERT J.
; REGISTRATION NUMBER: 33,915
; REFERENCE/DOCKET NUMBER: 1227-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 402-398-9000
; TELEFAX: 402-398-9005
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 324 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-671-320-11
```

Query Match 37.0%; Score 594.5; DB 2: Length 324;  
Best Local Similarity 89.9%; Pred. No. 4.5e-57;  
Matches 123; Conservative 63; Mismatches 115; Indels 7; Gaps 5;

[illegible]

```

1      RESULT      4
2      US-08-868-577-11
3      ; Sequence 11, Application US/08868577
4      ; Patent No. 5866695
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Vierling Jr., Richard A
7      ; TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN
8      ; TITLE OF INVENTION: ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY
9      ; NUMBER OF SEQUENCES: 19
10     ; CORRESPONDENCE ADDRESSES:
11     ; ADDRESSEE: Rothwell, Figg, Ernst & Kurz
12     ; STREET: 555 13th Street NW, Suite 701 East
13     ; CITY: Washington
14     ; STATE: DC
15     ; COUNTRY: USA
16     ; ZIP: 20004
17     ; COMPUTER READABLE FORM:
18     ; MEDIUM TYPE: Floppy disk
19     ; COMPUTER: IBM PC compatible
20     ; OPERATING SYSTEM: PC-DOS/MS-DOS
21     ; SOFTWARE: WordPerfect 6.1
22     ; CURRENT APPLICATION DATA:
23     ; APPLICATION NUMBER: US/08/868,577
24     ; FILING DATE: 04-JUN-1997
25     ; CLASSIFICATION: 536
26     ; ATTORNEY/AGENT INFORMATION:
27     ; NAME: Jondle, Robert J.
28     ; REGISTRATION NUMBER: 33,915
29     ; REFERENCE/DOCKET NUMBER: N1227-003
30     ; TELECOMMUNICATION INFORMATION:
31     ; TELEPHONE: 402-333-1550
32     ; TELEFAX: 402-333-1510
33     ; INFORMATION FOR SEQ ID NO: 11:
34     ; SEQUENCE CHARACTERISTICS:
35     ; LENGTH: 324 amino acids
36     ; TYPE: amino acid
37     ; TOPOLOGY: linear
38     ; MOLECULE TYPE: protein
39     ; US-08-868-577-11

```

	Query Match	37.0%	Score 594.5;	DB 2;	Length 324;
	Best Local Similarity	39.9%;	Pred. No.4.5e-57;		
	Matches 123; Conservative	63;	Indels 115;	Gaps 5;	
QY	2 QUTPFYDSCNVSNIYRDTIVNELRSPPRIASILRLHFHDFCVNGCDASILDNTTS	61			
	: : :   :  :::           ::   :				

Db 22 QLOGLGFVANSCKPKAEQYLKRVHDIHHNAPSLLAALLRMHFHDFCVAGCDSAVLNMSTN 81

QY 62 FRTKEADAFGNANSARGPEYIDRMKAAYBSACPTVSCADLTTTAAQOSVTLAGCPMKRP 121

Db 82 -QAEKNAPPNL-TVRGPEFDIRKISLVEAECPGVASCADILTTLAARDPTIATGCGPEMKVP 139

QY 122 LGRRSLOAFDILNANLPAPEFPLPOLKDSFRVWGNRSDDLVALSGGHTFGKNOCRF 181

Db 140 TORRGGVSNLTLEARKNNITPAPSSNFTLTQTLFANQGLD-LKDLVLLSGATTIGIARCSL 198

QY 182 MDRLVNSNTGLPDPPTLNTTYTLQTLRGL-CPLNGNL-SALVDPDLRTPTTFDNKKYVNE 239

Db 199 SNRLTFNFGKGDOPPSLDSEFYAAMLKAFKCTDLNKLMTTIEMDPGSRKTFEJLSYSHVI 258

QY 240 EOKGLIQSDQELFSSPDATDPIPLVRSFANSTQFPFAFEAMDRMNITPLTGQOQR 299

Db 259 KRKGLESDALLT-NSVTTAAQITIIQLLESVEENFADFATSIEMKGRINVKTGTBEIR 316

QY 300 LNCRVVNS 307

Db .317 KHCAPINS 324

```

1      RESULT      5
2      US-08-671-1320-13
3      ; Sequence 13, Application US/08671320
4      ; Patent No. 5840558
5      ; GENERAL INFORMATION:
6      ; APPLICANT: VERLING JR, RICHARD A
7      ; TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN
8      ; TITLE OF INVENTION: ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY
9      ; NUMBER OF SEQUENCES: 17
10     ; CORRESPONDENCE ADDRESS:
11     ; ADDRESSEE: 1125 SO. 103RD STREET
12     ; STREET: SUITE 330
13     ; CITY: OMAHA
14     ; STATE: NE
15     ; COUNTRY: US
16     ; ZIP: 68124-1076
17     ; COMPUTER READABLE FORM:
18     ; MEDIUM TYPE: Floppy disk
19     ; COMPUTER: IBM PC compatible
20     ; OPERATING SYSTEM: PC-DOS/MS-DOS
21     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
22     ; CURRENT APPLICATION DATA:
23     ; APPLICATION NUMBER: US/08/671,320
24     ; FILING DATE:
25     ; CLASSIFICATION: 435
26     ; ATTORNEY/AGENT INFORMATION:
27     ; NAME: JONDEL, ROBERT J.
28     ; REGISTRATION NUMBER: 33,915
29     ; REFERENCE/DOCKET NUMBER: 1227-001
30     ; TELECOMMUNICATION INFORMATION:
31     ; TELEPHONE: 402-398-9000
32     ; TELEFAX: 402-398-9005
33     ; INFORMATION FOR SEQ ID NO: 13:
34     ; SEQUENCE CHARACTERISTICS:
35     ; LENGTH: 324 amino acids
36     ; TYPE: amino acid
37     ; TOPOLOGY: linear
38     ; MOLECULE TYPE: protein
39     ;
40     ; US-08-671-320-13

```

Query Match	36.8%	Score 591.5	DB 2:	Length 324:
Best Local Similarity	39.9%	Pred No. 5e-57:		
Matches 123:	Conservative 62:	Mismatches 110:	Indels 7:	Gaps 5:
QY	2 QLTPTTFYDNCSPVSNIVRTITVNELRSDDPIASTLRHFDHCVCNGDASILLDNFTTS	61		
DB	22 QQLGFGYAKSCPNADIVLKFVHDHIIHNAPSLAALIRHMFDFCEVRGCDASVLLINSTPN	81		
QY	62 PFTKDAFGNANASRGFPVVIDMRKAIVESACPRTPSCADLLITIAQOQSVTLAAGDGSWRVP	121		

Db	82	-QA&KN&P&PNL-TV&RG&E&D&I&R&I&SL&VE&A&E&C&P&G&V&SC&AD&I&LL&TS&ARD&T&IA&T&G&P&E&K&VP	139
Oy	122	LG&RD&SL&O&A&F&LD&N&AN&L&P&A&F&F&LP&OL&K&S&F&R&V&G&N&R&S&D&L&V&A&L&S&G&T&F&G&N&O&R&F&I	181
Db	140	T&GR&D&G&V&S&N&L&E&A&R&D&N&I&P&A&S&S&F&E&T&L&O&T&L&F&A&N&O&G&L&D-L&K&D&L&V&L&S&G&A&H&T&G&I&A&C&S&L	198
Oy	182	M&D&R&L&Y&N&S&N&T&G&L&P&P&T&L&N&T&Y&L&O&T&L&R&G&L--C&P&L&N&G&N&L&S&A&V&D&P&D&L&R&T&P&T&F&D&N&K&Y&V&N&L&E	239
Db	199	S&N&R&F&E&N&T&G&K&D&O&P&S&L&D&S&E&Y&A&N&L&K&A&F&C&T&D&L&K&M&T&T&I&E&M&P&G&S&R&K&T&F&D&L&S&Y&S&H&V	258
Oy	240	E&O&K&L&I&O&S&D&E&L&F&S&P&A&T&D&I&P&L&V&R&S&A&N&S&T&O&F&E&F&A&F&E&A&M&O&R&M&N&I&T&P&L&T&O&Q&O&I&R	299
Db	259	K&R&G&L&F&E&S&D&A&L&L&L--N&S&V&K&A&Q&I&I&E&L&B&S&V&E&N&F&A&F&A&T&S&M&E&K&M&R&I&N&V&K&T&O&T&B&E&I&R	316
Oy	300	L&N&C&R&Y&N&S	307
Db	317	K&H&C&A&F&L&N&S	324

```

: RESULT 6
: US-08-868-577-13
: Sequence 13, Application US/0886577
: Patent No. 5866695
: GENERAL INFORMATION:
: APPLICANT: Vierling Jr., Richard A
: TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN
: TITLE OF INVENTION: ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Rothwell, Figg, Ernst & Kurz
: STREET: 555 15th Street NW, Suite 701 East
: City: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Wordperfect 6.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/868,577
: FILING DATE: 04-JUN-1997
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Jondle, Robert J.
: REGISTRATION NUMBER: 33,915
: REFERENCE/DOCKET NUMBER: N1227-003
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 402-333-1550
: TELEFAX: 402-333-1510
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 324 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-868-577-13

```

Query Match	36.8%;	Score 591.5;	DB 2;	Length 324;
Best Local Similarity	39.9%;	Pred. No. 9.5e-57;		
Matches 123; Conservative	62;	Mismatches 116;	Indels 7;	Gaps 5

Qy 2 QLEPTEFEDNSCPVNSINVTAPDTIVLELSESDRIASILRLFEHOCFVNGCASILIDWTS 61  
Db 22 QLOQGEFAKSCPNMAEQIVLEKPVHDIHINAPSLAAALIRMHFHOCFVNGCASVILNSTN 81  
Qy 62 FRTKQDAFGNANSKRPVYIDRKKAAVESACPRVSCADLITLTAQDSVTLACGSPSRVP 121  
Db 82 -QAKRNASPPNLI-TVIRGDFEIDRLKSLVEACCPGVSCADLITLTSARQTIATGSGPEFWK 139

```
QY      122   LGRDYSIOAFJLDLNNANU|PAPEFFULPOLKOSFRVNGRSDI|VALSGGTHPCKNCRFI    161  
          |||| : | : ||||  
Db      140   TGRDDGYASNLTERKDNI|PAPSSNFTLÖTLFANOGID -LKDLVLSGAHGTIGAHCSSL    198  
  
QY      182   MDRLYNFSNTG|PDPTLNTTYLOTLRG| -CPLNGNISALVDPELDRTPTIFDNKKYYVNLE    239  
          |||| : | : ||||  
Db      199   SNRLFNTGKGDDPSDSEYAANKAFKC|DLNKLMTXRIEMDPGRKTFDLSYSHVI    258  
  
QY      240   EQKGLISDOEFLSPSPATDTIPIVRSFANS|TOSTEFFNAFYEAAMDRCGNITPLLTGOCOIR    299  
          ::||| : | : ::|||  
Db      259   KRRLFEESDALLLT -NSVTKAQILIELLEGSEVENFEAEFAITSMEKGMRIWKTGETGEIR    316  
  
QY      300   LNCRVYNS    307  
          : | : |||  
Db      317   KHCAFLNS    324
```

```

RESULT 7
US-08-671-320-15
; Sequence 15, Application US/08671320
; Patent NO. 5840558
; GENERAL INFORMATION:
; APPLICANT: VIERLING JR, RICHARD A
; TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN
; TITLE OF INVENTION: ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 1125 SO. 103RD STREET
; STREET: SUITE 330
; CITY: OMAHA
; STATE: NE
; COUNTRY: US
; ZIP: 68124-1076
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/671.320
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JONDLE, ROBERT J.
; REGISTRATION NUMBER: 33,915
; REFERENCE/DOCKET NUMBER: 1227-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 402-398-9000
; TELEFAX: 402-398-9005
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 313 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-671-320-15

```

Query Match	35.2%;	Score 566;	DB 2;	Length 313;
Best Local Similarity	44.4%;	Pred. NO. 5.7e-54;		
Matches 127;	Conservative 44;	Mismatches 107;	Indels 8;	Gaps 6

[illegible]



```
GENERAL INFORMATION:
APPLICANT: VIERLING JR., Richard A
TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN
TITLE OF INVENTION: ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg, Ernst & Kurz
STREET: 555 13th Street NW, Suite 701 East
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/868,577
FILING DATE: 04-JUN-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Jondle, Robert J.
REGISTRATION/DOCKET NUMBER: 33,915
REFERENCE/DOCKET NUMBER: N1227-003
TELEPHONE: 402-333-1550
TELEFAX: 402-333-1510
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 313 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-868-577-17

Query Match          35.1%; Score 565; DB 2; Length 313;
Best Local Similarity 43.6%; Pred. No. 7.3e-54;
Matches 125; Conservative 47; Mismatches 107; Indels 8; Gaps 6;

QY 3 LTPFVYNSCPNYSNIVRDTIVNELRSDPRIASILRLHFHDCFVNCDSILLDNTTSF 62
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 23 LSNLYSKTPEDECTVAKAVIDATARDKTVPAALLRMHFDVRCGASVLLNSGNS 82
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 63 RTEKDAFGNANSARGFVIDRMKAVESACPRVYSCADLLTTIAAQSVTLAGPSWVPL 122
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 83 KAEDGPVNV-SLHAEVVIDAAKKALEASCPGVVSCADLLAARDAVFLSGPTWDEPK 141
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 GRDSDLAFLDLANANLPAPFTLLPOLKDSFRVNGLNRRSSDLVALSGGHTFGKNQCFIM 182
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 142 GRKDGRTSKAS-ETROLPAPTFNLSQLRQSFSQRLS-GEGLVALSGGHTLGFSSCSFK 199
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 DRLYNFSNGLPDPPTLNTWYLGTLRGICPL-NGNLSALVDFDLRTPIEDNKYYVNLREQ 241
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 200 NRHNHNRATHEDEPSLPSTATKLTICPLKNAKNAKNGTSMDSST-TTFNPTFYRLILQQ 258
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 242 KGLIOSDOELFSSPDATDTIPLVRSFANSTOTFFNAFVEADMRGNI 288
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 KGLFSSQVVLNDP---DTKNLVAKATSKAFYDAFAKSMIMMSI 302
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-08-190-029A-12
; Sequence 12, Application US/08190029A
; Patent No. 5736363
; GENERAL INFORMATION:
; APPLICANT: EDWARDS, Richard Mark
; APPLICANT: BAWDEN, Lindsey
; TITLE OF INVENTION: IGF-II ANALOGUES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ALLEGRETTI & WITCOFF, LTD.
; STREET: 10 S. WACKER DRIVE, SUITE 3000
; CITY: CHICAGO
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60606
; COMPUTER READABLE FORM:
```

```
STREET: 10 S. WACKER DRIVE, SUITE 3000
CITY: CHICAGO
STATE: ILLINOIS
COUNTRY: U.S.A.
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/190,029A
FILING DATE: 28-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB92/01389
FILING DATE: 27-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9202401.7
FILING DATE: 05-FEB-1992
APPLICATION NUMBER: GB 9116325.3
FILING DATE: 29-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: JOHN J. McDONNELL
REGISTRATION/DOCKET NUMBER: 26,949
REFERENCE/DOCKET NUMBER: 94,062
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 121 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-190-029A-12

Query Match          18.6%; Score 298.5; DB 1; Length 121;
Best Local Similarity 56.2%; Pred. No. 2.9e-25;
Matches 68; Conservative 6; Mismatches 26; Indels 21; Gaps 3;

QY 1 MQLTFEYDNCNYSNIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLDNTT 60
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MQLTFEYDNCNYSNIVRDTIVNELRSDPRIAASILRLHFHDCVNGCDASIM----- 54
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 SEFTEDAFGN-----ANSARGF-----PVIDRMKAVESACPRVYSCADLLTTA 105
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 AYRSEFTICGGLVDTLQFVCGDGRGFEFSRPAASVSRSGIVEECFCFSCDLALLETYC 114
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 106 A 106
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 A 115
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-08-462-695-12
; Sequence 12, Application US/08462695
; Patent No. 5854025
; GENERAL INFORMATION:
; APPLICANT: EDWARDS, Richard Mark
; APPLICANT: BAWDEN, Lindsey
; TITLE OF INVENTION: IGF-II ANALOGUES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER & ALLEGRETTI, LTD.
; STREET: 10 S. WACKER DRIVE, SUITE 3000
; CITY: CHICAGO
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60606
; COMPUTER READABLE FORM:
```



MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,695  
FILING DATE: 5-JUN-1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/190,029  
FILING DATE: 28-FEB-1994  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB92/01389  
FILING DATE: 27-JUL-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: GB 9202401.7  
FILING DATE: 05-FEB-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: GB 9116325.3  
FILING DATE: 29-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: JOHN J. McDONNELL  
REGISTRATION NUMBER: 26,949  
REFERENCE/DOCKET NUMBER: 94,062-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
INFORMATION FOR SEQ. ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 121 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-462-695-12

Query Match 18.6%; Score 298.5; DB 2; Length 121;  
Best Local Similarity 56.2%; Pred. No. 2.9e-25;  
Matches 68; Conservative 6; Mismatches 26; Indels 21; Gaps 3;  
QY 1 MQLPTTYDMSCPNVSIVIVDTIVNELRSDPRIAASLRLHFHDCFPVNGCDASILLDNTT 60  
DB 1 MQLPTTYDMSCPNVSIVIVDTIVNELRSDPRIAASLRLHFHDCFPVNGCDASM----- 54  
QY 61 SPFRKDAFGN-----ANSARGF-----PYIDRMKAAYESACPRIVSCADLLTIA 105  
DB 55 AYRPSEILCGELVDTLQFYVCGDRGFYRSPASRVSRSRSGIYECCFRCDLALLETYC 114  
QY 106 A 106  
DB 115 A 115  
RESULT 13  
US-08-313-185-49  
Sequence 49, Application US/08313185  
Patent No. 5851763  
GENERAL INFORMATION:  
APPLICANT: Heym, Beate  
APPLICANT: Cole, Stewart  
APPLICANT: Young, Douglas  
APPLICANT: Zhang, Ying  
APPLICANT: Honore, Nadine  
APPLICANT: Telenti, Amalio  
TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance  
TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance  
NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Parabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington

STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/313,185  
FILING DATE: 12-OCT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 02356.0068-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ. ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 726 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-313-185-49

Query Match 7.4%; Score 118.5; DB 2; Length 726;  
Best Local Similarity 23.1%; Pred. No. 0.00029;  
Matches 61; Conservative 24; Mismatches 88; Indels 91; Gaps 9;  
QY 96 VSCADLTITAAQSVTLAGPSRVRPLGRDLSQAFID-----LANAN 138  
DB 157 ISWADLEFLAGNALENGSRTEFGAGRDVMPDLDVNMGEKAWLTRHREALAKAP 216  
QY 139 LPA-----PFTTLPOLKDSFNNVGLNRSDDLVALSGGHTFGKNCRF 180  
DB 217 LGATEMGLIYVNEGPSPHSEPLSAAAIATRTGNNMGNDERYALAGGHTLAKTH--- 273  
QY 181 IMRLVNSMTGLPDPPTLNTYTLQTLGCLPLNGN-----LSALVDFDLRPPITDNK 233  
DB 274 ---GAPPTSNVG-PDPAAPLIEQGL-GWASTYSGVGADAIITSGLVWMTPTQMSNY 328  
QY 234 YYVNL-----EKGGLIOSDOELFSSPDATDTIP----- 262  
DB 329 FFENLFKYEWWQTRSPAGAIO-----FEAVDAPPEIIPDPSPKRRKPTMLVTDLRFD 383  
QY 263 ----LVRSPFANSOTPFNAFVEA 281  
DB 384 PEFEKISRFLNDPQAFNEAFARA 407  
RESULT 14  
US-08-459-499-13  
Sequence 13, Application US/08459499  
Patent No. 5871912  
GENERAL INFORMATION:  
APPLICANT: Heym, Beate  
APPLICANT: Cole, Stewart T.  
APPLICANT: Young, Douglas B.  
APPLICANT: Zhang, Ying  
TITLE OF INVENTION: Nucleic Acid Probes, Sequences, and Methods  
TITLE OF INVENTION: for Detecting Mycobacterium Tuberculosis Resistant to Isoni  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Parabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: DC

```

1 COUNTRY: USA
2 ZIP: 20005-3315
3
4 COMPUTER READABLE FORM:
5
6 MEDIUM TYPE: Floppy disk
7
8 COMPUTER: IBM PC compatible
9
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11
12 SOFTWARE: Patentin Release #1.0, Version
13
14 CURRENT APPLICATION DATA:
15
16 APPLICATION NUMBER: US/08/459,499
17
18 FILING DATE: 02-JUN-1995
19
20 CLASSIFICATION: 536
21
22 PRIOR APPLICATION DATA:
23
24 APPLICATION NUMBER: US 07/875,940
25
26 FILING DATE: 30-APR-1992
27
28 PRIOR APPLICATION DATA:
29
30 APPLICATION NUMBER: US 07/929,206
31
32 FILING DATE: 27-MAY-1992
33
34 PRIOR APPLICATION DATA:
35
36 APPLICATION NUMBER: US 08/029,655
37
38 FILING DATE: 11-MAR-1993
39
40 ATTORNEY/AGENT INFORMATION:
41
42 NAME: Meyers, Kenneth J.
43
44 REGISTRATION NUMBER: 25,146
45
46 REFERENCE/DOCKET NUMBER: 03495.0110-03000
47
48 TELECOMMUNICATION INFORMATION:
49
50 TELEPHONE: 202-408-4000
51
52 TELEFAX: 202-408-4400
53
54 INFORMATION FOR SEQ ID NO: 13:
55
56 SEQUENCE CHARACTERISTICS:
57
58 LENGTH: 726 amino acids
59
60 TYPE: amino acid
61
62 TOPOLOGY: linear
63
64 MOLECULE TYPE: peptide
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969

```

Query Match	7.48;	Score 118.5;	DB 2,	Length 726;
Best Local Similarity	23.18;	Pred. No. 0.00029		
Matches	61;	Conservative	88;	Indels 91;
		Mismatches	88;	Gaps 9

```

QY      96  VSCADLLTIAAQSVTLTAGGSGMNPVLRPRDLSQAFID-----LANAN 138
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      157  ISMADFLFAGNVALLENSGFFTFGAGAGREDVWPEDDLVNMGDEKAKLTIRHREALAKAP 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      139  LPA-----PFTTLPOLKDSFRNVLGNRSSDLVALSGCHTFGKNOCRF 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      217  LGAETEMGLIYVNEPGDPHSGEPLISAATAIRATFCFNMGMNDEETVALIAGCHTLEKTH--- 273
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      181  IMDRLYNFSNTGLDPDTLTNTTYLTTLGLCPDLGN-----LSALVDPLRPTPIFDNK 233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      274  ---GAGPFSNNG-PDPEAAPTEEDGL-GMASTYGSYGADAITSGLEVWVTOTPTOMSNY 328
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      234  YYVNL-----EEOKGLIOSDOLFESSPDATDTIP-----262
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      329  FFENILFYEWQTRSPAGAID-----FEAVADAPEIIPDPEDPSKKRPMLVTDLTIRFD 383
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      263  -----LVRSFANSOTOTFFENAVEA 281
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      384  PEFEKISRRLNDPOAFNEAFARA 407

RESULT 15
US-08-617-697-10
: Sequence 10, Application US/08617697
: Patent No. 5977336
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

```

STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/617,697  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 05-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Beirstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-557  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1600 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

Query Match	7.2%;	Score 116;	DB 2;	Length 1600;
Best Local Similarity	23.0%;	Pred. No. 0.0019;		
Matches	82;	Conservative	45;	Mismatches 118;
				Indels 112;
				Gaps 18

OY	14	NVSNIVRPTIV-----	:	----	NELRSPIAASTILRHFDHC-----	FVNGCASI	54
Dd	661	NVSSTLNTVOKFTFIKFEVDGSGNODLRSSRFAGY---	:	----	HFGIGKTINENIGANAKA	717	
OY	55	LL-----DNTTSFTEEDAFGMANSANGFPVIDRKAAVESACPRVSCADLL	:	----		102	
Dd	718	LFLTKPNATDPKKELPITTFVANITTAAGNSDSVMFYI-	:	----	-HANLT	760	
OY	103	TIAA---OOSVTLAGCEPMWRPLGRDRSIQAF--LDLANANLPAPFFTLPOLKDFSRNV	:	----		156	
Dd	761	SRAIGINMDSINITIGIDFSITSNNRRS-NAFELIKDL-TINTAGSNFSLKOTFRDYNE	:	----		818	
OY	157	GLNRSSDLVALSGCHTFG-----KNOCFTIMORLYNFSNTGLPDPTL	:	----		198	
Dd	819	YSKHAINSSHMLTLLGGVYTLGGGENSSSITGTINITYNKANVYLQADATSNSNTGLKFRRL	:	----		878	
OY	199	NTFYLOTLRGCLPLNCNLALSALVDEDLRTPTFDNKKYYVNLEEOGKLQSQOELEFSPDAT	:	----		258	
Dd	879	-----TL-GNISVECNISLT-----GANANIYCNLSIAEDSFPGK-EAS	:	----		915	
OY	259	DTIPLVSEANSTOTFFNAPYEADMGNCT-----TPULGTGOQIHLNCRAYN	:	----		306	
Dd	916	DNLNLTGFETNNGTANTN--IKGVYVKLGDINNKGLNITTINASTOKTYT--INGNITN	:	----		969	

Search completed: October 4, 2000, 12:38:12  
Job time: 21429 sec

---

**THIS PAGE BLANK (USPTO)**



C:Accession: S00625; S32972; A00502  
R: Fujiyama, K.; Takemura, H.; Shibayama, S.; Kobayashi, K.; Choi, J.K.; Shimmyo, A.; Tak  
Eur. J. Biochem. 173, 681-687, 1988  
A:Title: Structure of the horseradish peroxidase isozyme C genes.  
A:Reference number: S00625; MUID:88225087  
A:Accession: S00625  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-353 <FUJ>  
R: Weinder, K.G.  
Eur. J. Biochem. 96, 483-502, 1979  
A:Title: Amino acid sequence studies of horseradish peroxidase. Amino and carboxyl termi  
radish peroxidase C. S32972; MUID:79236311  
A:Reference number: S32972; MUID:79236311  
A:Accession: S32972  
A:Molecule type: protein  
A:Residues: 31-338 <WEI>  
R: Weinder, K.G.  
FEBS Lett. 72, 19-23, 1976  
A:Title: Covalent structure of the glycoprotein horseradish peroxidase (EC 1.11.1.7).  
A:Reference number: A00502; MUID:77068850  
A:Accession: A00502  
A:Molecule type: protein  
A:Residues: 31-338 <WEI>  
C:Genetics:  
A:Gene: prxCl  
A:Superfamily: peroxidase  
C:Keywords: chromoprotein; glycoprotein; heme; iron; metalloprotein; oxidoreductase; py  
F:1-30/Domain: signal sequence #status predicted <SIG>  
F:31-338/Product: peroxidase C1 #status experimental <AT>  
F:31/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experim  
F:41-121,74-79,127-331,207-239/Disulfide bonds: #status experimental  
F:43,87,188,216,228,244,285,298/Binding site: carbohydrate (Asn) (covalent) #status exp  
F:68/Active site: Arg #status predicted  
F:72,200/Binding site: heme iron (His) (axial ligands) #status predicted

Query Match 99.4%; Score 1598; DB 1; Length 353;  
Best Local Similarity 99.7%; Pred. No. 4.5e-128;  
Matches 307; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QLTPTPTDNCSPVNSIVRTIVNELSDPRIASILRLHEDCVNGCDASILDNTTS 61  
DB 31 QLTPTPTDNCSPVNSIVRTIVNELSDPRIASILRLHEDCVNGCDASILDNTTS 90  
QY 62 FRTKFAFGNANSARGPVIDRMKAAYESACPRIVSCADLLTIAQGSYTLAGSPWRVP 121  
DB 91 FRTKFAFGNANSARGPVIDRMKAAYESACPRIVSCADLLTIAQGSYTLAGSPWRVP 150  
QY 122 LGRRDSIQAFDLANANLPAFFTLPLQKDSFRNVLGRSSDVLVLSGHTFGKNQCRPI 181  
DB 151 LGRRDSIQAFDLANANLPAFFTLPLQKDSFRNVLGRSSDVLVLSGHTFGKNQCRPI 210  
QY 182 MDRLVNFSGTGLPDPILNTTYLQTLRGLCLPLNGNLALVDLFDLRTPTTFDKNYVNLLEEQ 241  
DB 211 MDRLVNFSGTGLPDPILNTTYLQTLRGLCLPLNGNLALVDLFDLRTPTTFDKNYVNLLEEQ 270  
QY 242 KGLISDQDELFPSSPDATDTLPIVRSFANSNOTFFNAFEANDRMKNITPLTGTOGIRLN 301  
DB 271 KGLISDQDELFPSSPDATDTLPIVRSFANSNOTFFNAFEANDRMKNITPLTGTOGIRLN 330  
QY 302 CRVYVNSNS 309  
DB 331 CRVYVNSNS 338

RESULT 3  
T46118  
peroxidase - Arabidopsis thaliana  
N:Alternate names: protein T213.40  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 18-Feb-2000  
C:Accession: T46118

R.Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Mayer, submitted to the Protein Sequence database, November 1999  
A:Reference number: Z23023  
A:Accession: T46118  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-353 <RMB>  
A:Cross-references: EMBL:AL132967  
A:Experimental source: cultivar Columbia; BAC clone T2J13  
C:Genetics:  
A:Map position: 3  
A:Introns: 77/3; 141/3; 198/1  
A:Note: T2J13.40  
C:Superfamily: peroxidase

Query Match	93.8%	Score 1509;	DB 2;	Length 353;
Best Local Similarity	92.9%	Pred. No. 1.6e-120;		
Matches 286;	Conservative 15;	Mismatches 7;	Indels 0;	Gaps 0;
QY	2	QUTPEFYNSCENVNINIRDTIVNELRSDDPRIAASIIIRLHFHDFVNCDCASILLDNPTS	61	
DB	31	QUTPEFYRSCENVNINIRDTIVNELRSDDPRIAASIIIRLHFHDFVNCDCASILLDNPTS	90	
QY	62	FRTKEDAFGNANSARGPFVIDRMKAIVESACDRIVSCADLLITAAQSVTLAGBSWRVP	121	
DB	91	FRTKEDAFGNANSARGPFVIDRMKAIVERACPRTVSCADMLITAAQSVTLAGBSWRVP	150	
QY	122	IGRRSLAFLDILANANIPAPFTTLPOLKDSFRNGLRRSSDIALSGGHFPGKNOCRFI	181	
DB	151	IGRRSLAFLDILANANIPAPFTTLPOLKASFRNGLDRPSDIALSGGHFPGKNOCFI	210	
QY	182	MDRLNFSNTGLPDEPTLTNTTYLQTLRGCLPNGNLSALVDEDLRTPTTFDMKKVYNLEEQ	241	
DB	211	LDRLNFSNTGLPDEPTLTNTTYLQTLRGCLPNGNSALVDEDLRTPTTFDMKKVYNLEEK	270	
QY	242	KGLIOSDELEFSSPATTTIPLVNSFANSTOTFFPAFVPEANDRMGNITPLTGTGQIRLN	301	
DB	271	KGLIOSDELEFSSPATTTIPLVRAVADGTQTFEVAFVEMRMGNITPTGTGQIRLN	330	
QY	302	CRVNSNS	309	
DB	331	CRVNSNS	338	
RESULT	4			
S37495				
peroxidase (EC 1.11.1.7) Cb - Arabidopsis thaliana				
C.Species: Arabidopsis thaliana (mouse-ear cress)				
C.Date: 09-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 04-Mar-2000				
C.Accession: S37495				
R.Intlapruk, C.; Takano, M.; Shimmyo, A.				
submitted to the EMBL Data Library, April 1993				
A.Description: Nucleotide sequence of a new cDNA for peroxidase from Arabidopsis thal				
A.Reference number: S37495				
A.Accession: S37495				
A.Molecule type: mRNA				
A.Residues: 1-353 <INT>				
A.Cross-references: EMBL:X71794; NID:9405610; PIDN:CAA50677.1; PID:9405611				
C.Genetics:				
A.Gene: prxhc				
C.Superfamily: peroxidase				
C.Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase				
F:41-121/Disulfide bonds: #status predicted				
F:68/Active site: Arg #status predicted				
F:72,200/Binding site: heme iron (His) (axial ligands) #status predicted				
F:74-79/Disulfide bonds: #status predicted				
F:127-131/Disulfide bonds: #status predicted				
F:207-239/Disulfide bonds: #status predicted				
Query Match	93.3%	Score 1500;	DB 2;	Length 353;
Best Local Similarity	92.2%	Pred. No. 9.3e-120;		

```

Matches 284; Conservative 15; Mismatches 9; Indels 0; Gaps 0;
QY 2 QLTPEFYDNCSPVNSVNIKRTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDNNTS 61
Db 31 QLTPEFYDRCSPVNTIVIRETIVNELRSDPRIASILRLHFHDCFVNGCDASILLDNNTS 90
QY 62 FRTEKDAFGNANSARGPFVIDIRMAKAAVESACPRIYSCADLLTTIAAQSVTLAGPSMRVP 121
Db 91 FRTEKDRFGNANSARGPFVIDIRMAKAAVERACPRIVSCADMJTIAAQSVTLAGPSMRVP 150
QY 122 LGRRDSLOAFIDLNNANLPAPFFTLPLKDSFRRVNGLRSSDVALSGHFGKNOCFI 181
Db 151 LGRRDSLOAFLELANANLPAPFFTLPLKASFRVNGLDPRSDVALSGHFGKNOCFI 210
QY 182 MDRLNFSNTGLPDPPTLNTTYTLQTLRLGLCPPLNGNLSALVDEDLRTPTLFDNKKYVNLREQ 241
Db 211 LDRFNFSNTGLPDPPTLNTTYTLQTLRLGLCPPLNGRNSALVDEDLRTPTLFDNKKYVNLKER 270
QY 242 KGLIOSDOELFSSPDATDTTIPLVNSFANSTOTFFNAPVEANDRMGNITPLTGTGQIRLN 301
Db 271 KGLIOSDOELFSSPDATDTTIPLVNAYADGTOTFFNAPVEANRMGNITPTGTGQIRLN 330
QY 302 CRVYNSNS 309
Db 331 CRVYNSNS 338

RESULT 5
J00457
peroxidase (EC 1.11.1.7) C - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 04-Mar-2000
C:Accession: J00457; T46119
R:Imapuk, C.; Higashimura, N.; Yamamoto, K.; Okada, N.; Shimmyo, A.; Takano, M.
Gene 98, 237-241, 1991
A:Title: Nucleotide sequences of two genomic DNAs encoding peroxidase of Arabidopsis thaliana
A:Reference number: J00457; M01D:91200671
A:Accession: J00457
A:Molecule type: DNA
A:Residues: 1-354 <INM>
A:Cross-references: GB:58380; NID:g166826; P10N:AAA32849.1; PID:g166827
R:Rieger, M.; Gabai, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Mayer, K.
Submitted to the Protein Sequence Database, November 1999
A:Reference number: Z23023
A:Accession: T46119
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-354 <RIE>
A:Cross-references: EMBL:AL132967
A:Experimental source: cultivar Columbia; BAC clone T2J13
C:Genetics:
A:Gene: prxCa; T2J13.50
A:Map position: 3
A:Introns: 78/3; 142/3; 199/1
C:Superfamily: peroxidase
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:4-122/Disulfide bonds: #status predicted
F:69/Active site: Arg #status predicted
F:73/201/Binding site: heme iron (His) (axial ligands) #status predicted
F:75-80/Disulfide bonds: #status predicted
F:128-332/Disulfide bonds: #status predicted
F:208-240/Disulfide bonds: #status predicted
Query Match 91.7%; Score 1475; DB 2; Length 354;
Best Local Similarity 90.9%; Pred. No. 1.2e-117;
Matches 280; Conservative 13; Mismatches 15; Indels 0; Gaps 0;
QY 2 QLTPEFYDNCSPVNSVNIKRTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDNNTS 61
Db 32 QLTPEFYDRCSPVNTIVIRETIVNELRSDPRIAGSILRLHFHDCFVNGCDASILLDNNTS 91
QY 62 FRTEKDAFGNANSARGPFVIDIRMAKAAVESACPRIYSCADLLTTIAAQSVTLAGPSMRVP 121

```

Query Match	91.6%	Score 1473;	DB 2;	Length 332;
Best Local Similarity	90.9%	Pred. No. 1.7e-117;		
Matches 280;	Conservative 13;	Mismatches 15;	Indels 0;	Gaps 0;
2 QLTPTFYDSCPNVSNIVRDTIVNELKSDPRIAASIRLHFHOCFVNGCDASILLDDMTS 61				
10 QLTPTFYDSCPNVSNIVRDTIVNELKSDPRIAASIRLHFHOCFVNGCDASILLDDMTS 69				
62 FTEKDAFGNANSARCEPVVIDRKKAAVESACPTVSCADLLTTAAQOSVTLAGPSWRVP 121				
70 FTEKDAFGNANSARCEPVVIDRKKAAVESACPTVSCADLLTTAAQOSVTLAGPSWRVP 129				
122 LGRRDSLOAFLLDIANNLPAFFTLPOLKNSFRVNGINRSSDLVALSGGHFGKNOCRFI 181				
130 LGRRDSLOAFLLDIANNLPAFFTLPOLKNSFRVNGINRSSDLVALSGGHFGKNOCRFI 189				
182 MDRLVNFSTGGLPDTLNTTYYLQTLRGICPLNGNLSTALVPDFDLTPTIFDKKYYVNLLEQ 241				
190 MDRLVNFSTGGLPDTLNTTYYLQTLRGICPLNGNLSTALVPDFDLTPTIFDKKYYVNLLEQ 249				
242 KGLIISDQELFSSPDATDTTPIVRSFANSTQTEFFNAFVEMADRMGNITPPLTGOGIIRLN 301				
250 KGLIISDQELFSSPDATDTTPIVRSFANSTQTEFFNAFVEMADRMGNITPPLTGOGIIRLN 309				





F;212,214/Binding site: substrate (Arg, Tyr) #status predicted

Query Match	70.0%;	Score 1125.5;	DB 2;	Length 349;
Best Local Similarity	70.9%;	Pred. No. 5.4e-88;		
Matches 217;	Conservative 34;	Mismatches 54;	Indels 1;	Gaps 1;

Oy 2 QLTPEFFYDNCSPVSNIVRDTIVNELRSDPRIASLRLHHEHDFVNGCDASILDNWTS 61  
|||::|||:-|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db 30 QLRPDEYFRICPSVENIGDIIVDELRTDIRIASLRLHFHCFCFVRGCDASILDNSTS 89

```

QY      62  FRTKDAFGANSARKEFPVIDMKAAVESACPRVSCADLLITIAAQSIVTLAGGPSKRPV 120
          |||||  |||||  |||||  ::  |||||  |||||  :||  ||  ||  ||
Db      90  FRTKDAAPRANSARKEFVIDMKKTSLEKACPRVSCADVLTIASQISVLTSGGPMKRPV 149

```

**Oy** 122 LGRDLSQAFIDLANNANLPAPFFTLPLQLKDSFRVYGLNRRSDVALSGGHFGKNCRFI 181  
|||||::|| ||| :||||| ||| ::||| :||||| ||| ::||| :  
**Db** 150 LGRDSEAEFIDLANNALPSFFTLaOLKKAFADVGLNNRPSDLVALSGGHFGRAQQQEV 209

```

QY  182 MDRLYNESNGCLPDPPLNTTYYIQTLRGICPLNGLSALVDFDLRPTTIFDNKYYVYNLEEQ 241
      ||||: | ||||: || | ||||: ||: ||||: || | ||||: || || ||
DB  210 TPRLYNENGTRPDPPLDPTLYVLQRLALCPQNGTGVLVNEDVVPNTFDQYYTINLRNG 269

```

QY 242 KGLIQSDQDELFSSPDATDTIPLVRSFANSTQTFENAFVEMDRMGNITPLTGTQGRRLN 301  
|||||:| | | | | :| | | | | :| | | | | |  
db 270 KGLIQSDQDELFSTPGA-DTIPLVNLSSNTFAFFGAFVDMIRMGNLRLPTGTQGEIRQN 328

Qy	302	CRVYNS	307
Db	329	CRVYNS	334

RESULT 10  
T02506  
peroxidase (EC 1.11.1.7) T19C21.12 - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)

R-Rounsley, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kau submitted to the EMBL Data Library, August 1998

Aridiscripton: Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.

A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-349 <R0U>

C:Genetics: 2  
A:Map position: 2  
A:Introns: 76/3; 140/3; 197/1

C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase  
F:40-120/Disulfide bonds: #status predicted  
F:67/Active site: Arg #status predicted

```

F:73-78/Disulfide bonds: #status predicted
F:126-329/Disulfide bonds: #status predicted
F:206-238/Disulfide bonds: #status predicted

```

Query Match	68.9%	Score 1108.5	DB 2	Length 349
Best Local Similarity	70.5%	Pred. No. 1.5e-86		
Matches 215	Conservative 31	Mismatches 58	Indels 1	Gaps 1

```

30 QLRPFYFRICPIENIGDTIVNELRDPRIASLRLHFHDCFVRGCDASILLNSTS 89

```

```
Db      90 ERTEKDAPNNNSVRCFDVIDNKKAAIERACPRIVSCADITTIASQISVLSSGGPMWHPVP 149
```

QY 122 LGRDLSIAFLDLANANLPAPFFTLPLQKDSFRNVGLNRSSDLVALSGGHTFGKNQCRFI 181

Db 150 LGRHDSVEAFALANTALPSPSTLTQLKTAFAVGYLNRHSDVLALSGGHIFGKACQFV 209

Oy 182 MDRLYNFSNTGLPDPPTINTYYQLRLGLCP LINGHSALVDFDLRPPTIFDNKYYVNLTEEQ 241

Db 210 TPRLYNFGNTRRDPDLSNPYTVLVELRLRCPPQNGNGTVLNVFDSVPTTFDQGYTNLLNG 269

Qy 242 KGLIOSDQDELFSSPDATDTIPLVRSFANSTOTFENAFVEAMDRCNITPLTGQGIPLN 301

Db	270	KGLIGSDQVLESTPGADTILPVNQSSNTFVFPGAFVDAMIRMGSLKPLTGTGGELRQN	328
Qy	302	CRVYN	306

D5	329	CRWN	333
----	-----	------	-----

peroxidase (EC 1.11.1.7) T19C21.13 - *Arabidopsis thaliana*  
 10250/  
 C:species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 05-Mar-1999 #sequence revision 05-Mar-1999 #text change 04-Mar-2000

K  
S.M.J.  
Sykes,  
R.C.;  
Brannon,  
M.L.;  
Crosby,  
A.A.;  
Netchum,  
J.;  
Lin,  
D.;  
Kounsley,  
S.D.,  
submitted  
to EMBL  
Data Bank,  
August  
1998

A: Description: Arabidopsis thaliana chromosome II BAC T19c21 genomic sequence.

A: Reference number: Z14676

A:Cross-references: EMBL:AC004683; NID:g3395421; PIDN:AAC28766.1; PID:g3395434  
A:Residues: 1-349 <R0U>  
A:Molecule type: DNA  
A:Status: translated from g3, EMBO, UDB

A;Map position: 2  
A;Introns: 76/3; 140/3; 197/1  
A;Note: T19C21.13

F:71,199/Binding site: heme iron (His) #status predicted

F;126-329/Disulfide bonds:	#status predicted
F;206-238/Disulfide bonds: <td>#status predicted</td>	#status predicted

Query Match	67.18;	Score 1079.5;	DB 2;	Length 349;
Best Local Similarity	68.28;	Pred. NO. 4.3e-84;		
Matches 208; Conservative	37;	Mismatches 59;	Indels 1;	Gaps 1;

30 QLRPDEVFGTCFVFVDIIIGNIIVDELQTDPRIAASLRLHFHDCFCVRGCDASILLDNSTS 89

90 FTREKDAFPNNSARGFVIDRKKVALEERACPGRVSCADILTIASQISVLLSGGPMWPVP 149

Db	150	IGRRDSVEAFPALANALPSPEENLTQLKTAFAADVGLNRTSDLVALSGGHFGRAQCQFV	209
Qy	182	MDRLNFSNFGLPDPTLNTTYLTQLRGICPLNGNLSALVDFDLKRTPTIFDNKYYVNIIEEQ	241

0y 242 KGLIOSDOELEFSSPDATDTIPLVRSFANSTQTFENAFVEAMDRCMNTPTLTGOGIRLN 30|||||:| | |||| :| | |:| | | | | | | | |

QY 302 CRVN 306





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 13:04:55 ; Search time 27.02 Seconds

(Without alignments)  
354.391 Million cell updates/sec

Title: US-09-246-451-17

Perfect score: 1608

Sequence: 1 MQLPTFYDNCSPVNSIVR.....PLTGTGQQLRLNCRVNSNS 309

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1598	99.4	353	1 PERA_ARMRU	P00433 armoracia r
2	1475	91.7	354	1 PERC_ARATH	P24101 arabidopsis
3	1473	91.6	332	1 PERC_ARMRU	P15233 armoracia r
4	1469	91.4	351	1 PERB_ARMRU	P15232 armoracia r
5	1271	79.0	347	1 PER2_ARMRU	P17179 armoracia r
6	1125.5	70.0	349	1 PER3_ARMRU	P17160 armoracia r
7	1065.5	66.3	349	1 PERE_ARMRU	P24102 arabidopsis
8	875	54.4	305	1 PERX_ARMRU	P80679 armoracia r
9	786.5	48.9	324	1 PERX_TOBAC	P11965 nicotiana t
10	745.5	46.4	316	1 PER1_ARMRU	P19135 cucumis sal
11	737.5	45.9	292	1 PER2_ARMRU	P05855 triticum ae
12	729.5	45.4	312	1 PER1_ARMRU	P37835 oryza sativ
13	714	44.4	314	1 PER2_ARMRU	P27337 hordeum vul
14	707	44.0	315	1 PER1_ARMRU	P02200 nicotiana s
15	702	43.7	322	1 PERX_ARMRU	P00434 brassica ra
16	696.5	43.3	296	1 PERX_ARMRU	P22106 arachis hyp
17	618.5	38.5	330	1 PER2_ARMRU	P37834 oryza sativ
18	587.5	36.5	326	1 PER1_ARMRU	P15003 lycopersico
19	463	28.8	364	1 PER1_ARMRU	P12437 solanum tub
20	462	28.7	351	1 PERX_ARMRU	P15004 lycopersico
21	440	27.4	363	1 PERX_ARMRU	P16147 lupinus pol
22	297	18.5	158	1 PERX_ARMRU	P001548 hordeum vul
23	243.5	15.1	170	1 PER2_ARMRU	P15984 triticum ae
24	203	12.6	80	1 PERX_ARMRU	P48534 plasm sativ
25	164	10.2	249	1 PER1_ARMRU	P005431 arabidopsis
26	156	9.7	249	1 PER1_ARMRU	P13029 escherichia
27	118.5	7.4	726	1 PER1_ARMRU	P00431 saccharomyc
28	111.5	6.9	727	1 PER1_ARMRU	P17750 salmonella
29	111.5	6.9	727	1 PER1_ARMRU	P19136 phanerocoe
30	96.5	6.0	382	1 PER1_ARMRU	P01293 rana catesb
31	94.5	5.9	1496	1 PER1_ARMRU	P30996 clostridium
32	93.5	5.8	1274	1 PER1_ARMRU	P056326 treponema p
33	92	5.7	463	1 PER1_ARMRU	

34	92	5.7	1500	1	CPSM_HUMAN	P31327 homo sapien
35	90.5	5.6	244	1	FIMB_BORPE	P33409 borietella
36	90	5.6	581	1	FIRB_ADE05	P11818 human adeno
37	90	5.6	1500	1	CPSM_RAT	P07756 rattus norv
38	89.5	5.6	468	1	ALST_CANAL	P07457 candida alb
39	88	5.5	372	1	LIG6_PHACH	P50622 phanerocoe
40	88	5.5	2009	1	SEC7_YEAST	P11075 saccharomyc
41	87.5	5.4	361	1	LIG_PHLRA	P20010 phlebia rad
42	86.5	5.4	763	1	ECHA_PIG	P29554 sus scrofa
43	86	5.3	647	1	COAT_ADYG	P24029 aleutian m1
44	85.5	5.3	225	1	YEIP_ECOLI	P33028 escherichia
45	85.5	5.3	682	1	HTF4_HUMAN	P09081 homo sapien

## ALIGNMENTS

RESULT	1	PERA_ARMRU	STANDARD;	PRT;	353 AA.
ID	P00433;				
AC	21-JUL-1986 (Rel. 01, Created)				
DT	01-APR-1990 (Rel. 14, Last sequence update)				
DE	15-FEB-2000 (Rel. 39, Last annotation update)				
DE	PEROXIDASE C1A PRECURSOR (EC 1.11.1.7).				
OS	PRXCIA OR HPRC1.				
OC	Armoracia rusticana (Horsederish) (Armoracia lappathifolia).				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;				
OC	Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;				
OC	Brassicaceae; Armoracia.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 88225087.				
RA	Fujiyama K., Takemura H., Shibayama S., Kobayashi K., Choi J.K.,				
RA	Shimizu A., Takano M., Yamada Y., Okada H.;				
RT	"Structure of the horseradish peroxidase isozyme C genes.";				
RL	Eur. J. Biochem. 173:681-687(1988).				
RN	[2]				
RP	SEQUENCE OF 31-338.				
RX	MEDLINE: 77068850.				
RA	Wellinder K.G.;				
RT	"Covalent structure of the glycoprotein horseradish peroxidase (EC				
RL	1.11.1.7)."				
RN	FEBS Lett. 72:19-23(1976).				
RP	[3]				
RX	X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).				
RA	Gajhe M., Schuller D.J., Henriksen A., Smith A.T., Poulos T.L.;				
RT	"Crystal structure of horseradish peroxidase C at 2.15-A resolution.";				
RL	Nat. Struct. Biol. 4:1032-1038(1997).				
RN	[4]				
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).				
RX	MEDLINE: 98272679.				
RA	Henriksen A., Schuller D.J., Meno K., Wellinder K.G., Smith A.T.,				
RT	Gajhe M.;				
RL	"Structural interactions between horseradish peroxidase C and the				
RL	substrate benzhydroxamic acid determined by x-ray crystallography.";				
RL	Biochemistry 37:8054-8060(1998).				
RN	[5]				
RP	X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).				
RA	Meno K., White C.G., Smith A.T., Gajhe M.;				
RT	Submitted (DEC-1998) to the PDB data bank.				
RL	"FUNCTION. REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,				
CC	BIOSYNTHESIS AND DEGRADATION OF LIGNIN. DEFENSE RESPONSE TOWARD				
CC	WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE				
CC	DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.				
CC	-1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.				
CC	-1- SUBUNIT: MONOMER.				
CC	-1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.				
CC	-----				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M37156; AAA33377.1; ALT\_SEQ.

DR PIR: A00502; OPRHC.

DR PIR: S00625; S00625.

DR PDB: 1ATJ; 04-FEB-98.

DR PDB: 2ATJ; 28-JAN-98.

DR PDB: 3ATJ; 23-DEC-98.

DR PFAM: PF00141; peroxidase: 1.

DR PRINTS: PR00456; PEROXIDASE.

DR PRINTS: PR00461; PLPEROXIDASE.

DR PROSITE: PS00435; PEROXIDASE\_1; 1.

DR PROSITE: PS00436; PEROXIDASE\_2; 1.

KM Oxidoreductase; Glycoprotein; Peroxidase; Heme; Multigene family;

KW Signal; 3D-structure.

FT SIGNAL 1 30

FT CHAIN 31 338

FT PROPEP 339 353

FT MOD\_RES 31 31

FT ACT\_SITE 68 68

FT ACT\_SITE 72 72

FT ACT\_SITE 200 200

FT DISULFID 41 121

FT DISULFID 74 79

FT DISULFID 127 331

FT DISULFID 207 239

FT CARBOHYD 43 43

FT CARBOHYD 87 87

FT CARBOHYD 188 188

FT CARBOHYD 216 216

FT CARBOHYD 228 228

FT CARBOHYD 244 244

FT CARBOHYD 285 285

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

AC P24101;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE NEUTRAL PEROXIDASE C PRECURSOR (EC 1.11.1.7).  
GN PRXCA.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Rosidae; eurosid II; Brassicales;  
OC Brassicaceae; Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 91200671.  
RA Intapur C., Higashimura N., Yamamoto K., Okada N., Shimmyo A.,  
RA Takano M.;  
RT "Nucleotide sequences of two genomic DNAs encoding peroxidase of  
RT Arabidopsis thaliana.";  
RL Gene 98:237-241(1991).

CC -1- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,  
CC BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD  
CC WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE  
CC DEPENDENT ON EACH ISOZYME/ISOPORM IN EACH PLANT TISSUE.  
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) - OXIDIZED DONOR + 2 H(2)O.  
CC -1- COFACTOR: BINDS HEME.  
CC -1- TISSUE SPECIFICITY: ROOTS.  
CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M58360; AAA32849.1; -.

DR PIR: J00457; J00457.

DR HSSP: P00433; 2ATJ.

DR PFAM: PF00141; peroxidase: 1.

DR PRINTS: PR00456; PEROXIDASE.

DR PRINTS: PR00461; PLPEROXIDASE.

DR PROSITE: PS00435; PEROXIDASE\_1; 1.

DR PROSITE: PS00436; PEROXIDASE\_2; 1.

KM Oxidoreductase; Glycoprotein; Peroxidase; Heme; Multigene family;

KW Signal.

FT SIGNAL 1 21

FT CHAIN 22 354

FT ACT\_SITE 69 69

FT ACT\_SITE 73 73

FT ACT\_SITE 201 201

FT DISULFID 42 122

FT DISULFID 75 80

FT DISULFID 128 332

FT DISULFID 208 240

FT DISULFID 208 240

FT CARBOHYD 88 88

FT CARBOHYD 217 217

FT CARBOHYD 229 229

FT CARBOHYD 245 245

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

Query Match 91.7%; Score 1475; DB 1; Length 354;  
Best local Similarity 90.9%; Pred. No. 7.9e-117;  
Matches 280; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 2 QLTPTFYDSCPNVSNIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLDNNTS 61

DB 32 QLTPTFYDSCPTVTINIVRDTIVNELRSDPRIAGSILRLHFHDCFVNGCDASILLDNNTS 91

QY 62 FRTKEKAFGNANSARGFPVIDRKAVERACPRTVSCADMLTIAAOSVTLAAGPSWVRP 121

DB 92 FRTKEKALGNANSARGFPVIDRKAVERACPRTVSCADMLTIAAOSVTLAAGPSWVRP 151

FT	DISULFID	186	218	BY SIMILARITY.
FT	CARBOHYD	22	22	POTENTIAL.
FT	CARBOHYD	66	66	POTENTIAL.
FT	CARBOHYD	195	195	POTENTIAL.
FT	CARBOHYD	207	207	POTENTIAL.
FT	CARBOHYD	223	223	POTENTIAL.
FT	CARBOHYD	264	264	POTENTIAL.
SO	SEQUENCE	332 AA;	36548 MW;	1938A450D595DFBE CRC64;

  

Query Match	Best Local Similarity	91.6%;	Score 1473;	DB 1;	Length 332;				
Matches	280;	Conservative	13;	Mismatches	15;	Indels	0;	Gaps	0;
QY	2	QLPTEFYDNCSPVNSVNYRDTIVNELRSDPRIASILRLHFHDCFCVNGCDASILLDNMTS	61						
Db	10	QLPTEFYDNCSPVNSVNYRDTIIINELRSDPSIAASILRLHFHDCFCVNGCDASILLDNMTS	69						
QY	62	FRTKEDAGNANSARGPVYIDRMKAAYVSAAPRIVSCADLLTTIAQOSVTLTAGPSMRVP	121						
Db	70	FRTKEDAGNANSARGPVYIDRMKAAYVSAAPRIVSCADVLTTLIAQOSVNLTAGPSMRVP	129						
QY	122	LGRDLSLOAFEDLANANLPAFFTLPOLKDSFRVNGVLRSSDLYALSGHFGKNCRCFI	181						
Db	130	LGRDLSRQAFEDLANANLPAFSTLPELKAFAFVANGVLRSDLYALSGHFGKNCRCFI	189						
QY	182	MDRLYNESNTGLPDPPTLTYTLQTLRGICPLNGNLSALVDFDLPTPTIFDNKRYVNEEQ	241						
Db	190	MDRLYNESNTGLPDPPTLTYTLQTLRGQCPNGNSVLYVDFDLPTPTIFDNKRYVNEEQ	249						
QY	242	KGLIOSDDELFPSPDADVTIPLVRSFANSTOTFRNAYEAMDRCNGNTPLTGTGQIRLN	301						
Db	250	KGLIOSDDELFPSPNADVTIPLVRSYADGTOTFFNAFVANNRGNITPLTGTGQIRLN	309						
QY	302	CRVYNSNS	309						
Db	310	CRVYNSNS	317						

  

PERB_ARMRU	STANDARD;	PRT;	351 AA.
AC	P15232;		
DT	01-APR-1990 (Rel. 14, Created)		
DT	01-APR-1990 (Rel. 14, Last sequence update)		
DT	15-FEB-2000 (Rel. 39, Last annotation update)		
DE	PEROXIDASE C1B PRECURSOR (EC 1.11.1.7).		
GN	PRKC1B OR HPRC2.		
OC	Armoreria rusticana (Horsereadish) (Armoreria laphatfolia).		
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
OC	Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;		
OC	Brassicaceae; Armoracia.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE; 88225087.		
RA	Fujiyama K., Takemura H., Shibayama S., Kobayashi K., Choi J.K.,		
RA	Shinmoy A., Takano M., Yamada Y., Okada H.;		
RT	"Structure of the horseradish peroxidase isozyme C genes."		
RL	Eur. J. Biochem. 173:681-687(1988).		
CC	-I- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,		
CC	BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD		
CC	WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE		
CC	DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.		
CC	-I- CATALYTIC ACTIVITY: DONOR + H(2)O(2) -> OXIDIZED DONOR + 2 H(2)O.		
CC	-I- COFACTOR: BINDS HEME.		
CC	-I- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		

or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch).

CC EMBL; M37157; AAA33378.1; -.  
DR PIR; S00626; S00626.  
DR HSSP; P00433; 2ATU.  
DR PFAM; PF00141; peroxidase\_1.  
DR PRINTS; PR00458; peroxidase.  
DR PRINTS; PR00461; PLPEROXIDASE.  
DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
DR PROSITE; PS00436; PEROXIDASE\_2; 1.  
KM Oxidoreductase; Glycoprotein; Peroxidase; Heme; Multigene family;  
Signal.  
FT SIGNAL 1 28  
FT CHAIN 29 351  
FT MOD\_RES 29 29  
FT ACT\_SITE 66 66 PEROXIDASE C1B.  
FT ACT\_SITE 70 70 PYRROLIDONE CARBOXYLIC ACID (BY  
FT ACT\_SITE 198 198 SIMILARITY).  
FT DISULFID 39 119 DISTAL HISTIDINE (BY SIMILARITY).  
FT DISULFID 72 77 PROXIMAL HISTIDINE (HEME AXIAL LIGAND).  
FT DISULFID 125 329 BY SIMILARITY.  
FT DISULFID 205 237 BY SIMILARITY.  
FT CARBOHYD 41 41 POTENTIAL.  
FT CARBOHYD 85 85 POTENTIAL.  
FT CARBOHYD 214 214 POTENTIAL.  
FT CARBOHYD 226 226 POTENTIAL.  
FT CARBOHYD 242 242 POTENTIAL.  
FT CARBOHYD 283 283 POTENTIAL.  
SQ SEQUENCE 351 AA; 38645 MW; 7A8C606A3928950B CRC64;

Query Match 91.4%; Score 1469; DB 1; Length 351;  
Best local Similarity 90.6%; Pred. No. 2.5e-116;  
Matches 279; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

QY 2 QLTPTFFDNCSPVNSIVRDTYVELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 61  
DB 29 QLTPTFFDNCSPVNSIVRDTYVELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 88  
QY 62 FTEKDAFGNANSARGFVIDRMKAVERACPTVSCADLLTTAAQOSVTLAGGSPMRVP 121  
DB 89 FTEKDAFGNANSARGFVIDRMKAVERACPTVSCADLLTTAAQOSVTLAGGSPMRVP 148  
QY 122 LGRDLSQAFLLDANANLPAFFETLPQKDSFRVNGLNSSDVLVALSGHTFGKNOCRFI 181  
DB 149 LGRDLSQAFLLDANANLPAFFETLPQKDSFRVNGLNSSDVLVALSGHTFGKNOCRFI 208  
QY 182 MDRLYNFSNGLPDPPTLNTTYLQTLRGCLPLNGNLSALVDFDLRTPTIFDNKRYVNLREQ 241  
DB 209 MDRLYNFSNGLPDPPTLNTTYLQTLRGCLPLNGNLSALVDFDLRTPTIFDNKRYVNLREQ 268  
QY 242 KGLIOSDOELFSSPDATDTPIVRSFANSQTGFENAFVEMDMRGNTTPLTGTOGOIRLN 301  
DB 269 KGLIOSDOELFSSPDATDTPIVRSFANSQTGFENAFVEMDMRGNTTPLTGTOGOIRLN 328  
QY 302 CRVYNSNS 309  
DB 329 CRVYNSNS 336

RESULT 5  
PER2\_ARMRU  
ID PER2\_ARMRU STANDARD: PRT: 347 AA.  
AC P17179;  
DT 01-ANG-1990 (Rel. 15, Created)  
DT 01-ANG-1990 (Rel. 15, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PEROXIDASE C2 PRECURSOR (EC 1.11.1.7).  
GN PRXC2.  
OS Armoracia rusticana (Horseradish) (Armoracia lappacea).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;

CC Brassicaceae; Armoracia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90323613.  
RA Fujiyama K., Takemura H., Shimmyo A., Okada H., Takano M.:  
RT "Genomic DNA structure of two new horseradish-peroxidase-encoding  
genes";  
RL Gene 89:163-169(1990).  
CC -1- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,  
CC BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD  
CC WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE  
CC DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.  
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) -> OXIDIZED DONOR + 2 H(2)O.  
CC -1- COFACTOR: BINDS HEME.  
CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.lsb-sib.ch/announce/>  
CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

CC EMBL; D90115; BAA14143.1; -.  
DR PIR; JH0149; JH0149.  
DR HSSP; P00433; 2ATU.  
DR PFAM; PF00141; peroxidase\_1.  
DR PRINTS; PR00458; peroxidase.  
DR PRINTS; PR00461; PLPEROXIDASE.  
DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
DR PROSITE; PS00436; PEROXIDASE\_2; 1.  
KM Oxidoreductase; Glycoprotein; Peroxidase; Heme; Multigene family;  
Signal.  
FT SIGNAL 1 24  
FT CHAIN 25 347  
FT MOD\_RES 25 25  
FT ACT\_SITE 62 62 PEROXIDASE C2.  
FT ACT\_SITE 66 66 PYRROLIDONE CARBOXYLIC ACID (BY  
FT ACT\_SITE 194 194 SIMILARITY).  
FT DISULFID 35 115 DISTAL HISTIDINE (BY SIMILARITY).  
FT DISULFID 68 73 PROXIMAL HISTIDINE (HEME AXIAL LIGAND).  
FT DISULFID 121 325 BY SIMILARITY.  
FT DISULFID 201 233 BY SIMILARITY.  
FT CARBOHYD 81 81 POTENTIAL.  
FT CARBOHYD 210 210 POTENTIAL.  
FT CARBOHYD 238 238 POTENTIAL.  
SQ SEQUENCE 347 AA; 38035 MW; 3EE9A2CFDECBAA49A CRC64;

Query Match 79.0%; Score 1271; DB 1; Length 347;  
Best local Similarity 79.4%; Pred. No. 1.1e-99;  
Matches 243; Conservative 24; Mismatches 39; Indels 0; Gaps 0;

QY 2 QLTPTFFDNCSPVNSIVRDTYVELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 61  
DB 25 QLTPTFFDNCSPVNSIVRDTYVELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 84  
QY 62 FTEKDAFGNANSARGFVIDRMKAVERACPTVSCADLLTTAAQOSVTLAGGSPMRVP 121  
DB 85 FTEKDAFGNANSARGFVIDRMKAVERACPTVSCADLLTTAAQOSVTLAGGSPMRVP 144  
QY 122 LGRDLSQAFLLDANANLPAFFETLPQKDSFRVNGLNSSDVLVALSGHTFGKNOCRFI 181  
DB 145 LGRDLSQAFLLDANANLPAFFETLPQKDSFRVNGLNSSDVLVALSGHTFGKNOCRFI 204  
QY 182 MDRLYNFSNGLPDPPTLNTTYLQTLRGCLPLNGNLSALVDFDLRTPTIFDNKRYVNLREQ 241  
DB 205 MDRLYNFSNGLPDPPTLNTTYLQTLRGCLPLNGNLSALVDFDLRTPTIFDNKRYVNLREQ 264  
QY 242 KGLIOSDOELFSSPDATDTPIVRSFANSQTGFENAFVEMDMRGNTTPLTGTOGOIRLN 301  
DB 269 KGLIOSDOELFSSPDATDTPIVRSFANSQTGFENAFVEMDMRGNTTPLTGTOGOIRLN 328



Qy	2	QUTPTFDNSCPNPNSTNYRDTIVNELSDPRILASILRLRHDOCFVNCGDASILDNPTS	61
Dd	30	QLRDFEFCRCPSVFNNIIIGIIVDELRTDPRIIASLRLHFHDCFCVRGCDASILDNSTS	89
Qy	62	FRTKEDAFGNANSRGPFVIDIRMKAAVASACPRTYSCADLLITIAAQSVCVTLAGSMMRPV	121
Dd	90	FRTKEDAPRNANSARGREVIDRMKTSLERACPRYSADVLTITASISVLSSGGPMFVP	149
Qy	122	LGRDLSIOALFLDIANANLPAPFTLLPOLKDSFRVNGLRSSDLVALSGCHTFGKNCRFI	181
Dd	150	LGRDSEAEAFEDLANTLALPSPFTTLOAKRKAFADVGKLRPSDLVALSGCHTFEGRAQCFFV	209
Qy	182	MDRLYNSNTGLPPPTLNTTYTLQTLRGCLPLANGNISALVYDLDTPRTTFDKKYVNLEEQ	241
Dd	210	TPLRYNFNGTRRPPPTIDPYLVOLRALCPCPNGNGTVLVNDVYTTPNTEDEQYYTNLRNG	269
Qy	242	KGLIOSPOELSSPDATPTPIPLVSFSANSTOTFFNAFEARDMGRIINPLTGTOGOIRLN	301
Dd	270	KGLIOSDELFSTGCA-DTIPLVNLYSNTTAFFGAIFYDAIMRMGNLRPLGTGGEIION	328
Qy	302	CRYVNS	307
Dd	329	CRYVNS	334
RESULT	7		
PERE_ARATH	ID	STANDARD;	PRT; 349 AA.
AC	P24102:		
DT	01-MAR-1992 (Rel. 21, Created)		
DT	01-MAR-1992 (Rel. 21, Last sequence update)		
DT	01-OCT-1996 (Rel. 34, Last annotation update)		
DE	BASIC PEROXIDASE E PRECURSOR (EC 1.11.1.7).		
GN	PREEA.		
OC	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
OC	Magnoliophyta; eudicotyledons: Rosidae; eurosids II; Brassicales;		
OC	Brassicaceae; Arabidopsids.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 91200671.		
RA	Inapuk C., Higashimura N., Yamamoto K., Okada N., Shimmyo A.,		
RA	Takano M.;		
RT	"Nucleotide sequences of two genomic DNAs encoding peroxidase of		
RL	Arabidopsis thaliana.";		
Gene	98:237-241(1991)."		
CC	-1- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,		
CC	BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD		
CC	WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE		
CC	DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.		
CC	-1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) -> OXIDIZED DONOR + 2 H(2)O.		
CC	-1- COFACTOR: BINDS HEME.		
CC	-1- TISSUE SPECIFICITY: ROOTS.		
CC	-1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
CC	EMBL; M58381; AAA32842.1; .		
DR	PIR; J00458; J00458.		
DR	HSSP; P00433; IATU.		
DR	PEAM; PF00141; peroxidase; 1.		
DR	PRINTS; PR00458; PEROXIDASE.		
DR	PRINTS; PR00461; PUEROXIDASE.		
DR	PROSITE; PS00435; PEROXIDASE_1; 1.		
DR	PROSITE; PS00436; PEROXIDASE_2; 1.		
DM	Oxidoreductase; Glycoprotein; Peroxidase; Heme; Multigene family;		

Query Match	Best Local Similarity	Matches	Score	DB	Length
66.3%;	67.5%;	37;	1065.5;	1;	349;
Conservative	37;	Mismatches	61;	Indels	1;
Gaps					
1.1;					
Signal	1	19			
CHAIN	20	349			
ACT_SITE	67	67			
ACT_SITE	71	71			
ACT_SITE	199	199			
ACT_SITE	40	120			
DISULFID	73	78			
DISULFID	126	329			
DISULFID	206	238			
DISULFID	86	86			
CARBOHYD	173	173			
CARBOHYD	187	187			
CARBOHYD	217	217			
CARBOHYD	243	243			
SEQUENCE	349 AA;	38172 MM;	6E02C5758C36A8F	CR664;	
Query Match	66.3%;	Score	1065.5;	DB	1;
Best Local Similarity	67.5%;	Pred. No.	2e-82;	Length	349;
Matches	206;	Conservative	37;	Mismatches	61;
Indels	1;	Gaps			
1.1;					
Signal	1	19			
CHAIN	20	349			
ACT_SITE	67	67			
ACT_SITE	71	71			
ACT_SITE	199	199			
ACT_SITE	40	120			
DISULFID	73	78			
DISULFID	126	329			
DISULFID	206	238			
DISULFID	86	86			
CARBOHYD	173	173			
CARBOHYD	187	187			
CARBOHYD	217	217			
CARBOHYD	243	243			
SEQUENCE	349 AA;	38172 MM;	6E02C5758C36A8F	CR664;	
Query Match	66.3%;	Score	1065.5;	DB	1;
Best Local Similarity	67.5%;	Pred. No.	2e-82;	Length	349;
Matches	206;	Conservative	37;	Mismatches	61;
Indels	1;	Gaps			
1.1;					
Signal	1	19			
CHAIN	20	349			
ACT_SITE	67	67			
ACT_SITE	71	71			
ACT_SITE	199	199			
ACT_SITE	40	120			
DISULFID	73	78			
DISULFID	126	329			
DISULFID	206	238			
DISULFID	86	86			
CARBOHYD	173	173			
CARBOHYD	187	187			
CARBOHYD	217	217			
CARBOHYD	243	243			
SEQUENCE	349 AA;	38172 MM;	6E02C5758C36A8F	CR664;	
Query Match	66.3%;	Score	1065.5;	DB	1;
Best Local Similarity	67.5%;	Pred. No.	2e-82;	Length	349;
Matches	206;	Conservative	37;	Mismatches	61;
Indels	1;	Gaps			
1.1;					
Signal	1	19			
CHAIN	20	349			
ACT_SITE	67	67			
ACT_SITE	71	71			
ACT_SITE	199	199			
ACT_SITE	40	120			
DISULFID	73	78			
DISULFID	126	329			
DISULFID	206	238			
DISULFID	86	86			
CARBOHYD	173	173			
CARBOHYD	187	187			
CARBOHYD	217	217			
CARBOHYD	243	243			
SEQUENCE	349 AA;	38172 MM;	6E02C5758C36A8F	CR664;	
Query Match	66.3%;	Score	1065.5;	DB	1;
Best Local Similarity	67.5%;	Pred. No.	2e-82;	Length	349;
Matches	206;	Conservative	37;	Mismatches	61;
Indels	1;	Gaps			
1.1;					
Signal	1	19			
CHAIN	20	349			
ACT_SITE	67	67			
ACT_SITE	71	71			
ACT_SITE	199	199			
ACT_SITE	40	120			
DISULFID	73	78			
DISULFID	126	329			
DISULFID	206	238			
DISULFID	86	86			

Query Match	54.4%	Score 875	DB 1	Length 305
Best Local Similarity	55.4%	Pred. No. 1.7e-66		
Matches 170	Conservative 53	Mismatches 82	Indels	Gaps
QY	2	QLTFPYNSCPNNSIYRDTIVNELRSDPRASILRHFHDCVYNGCDASILLDNNTTS	61	
Db	1	QLMAFYSGTCPMNSAIVRSTIQAFQSDTRGASLIRHFHDCVDCDASILLDDSGS	60	
QY	62	FRTEDADGNNNSARGFVVIDIRMKAAVDSACPRVSCADLLTIAAQSVTLAGGSPWRP	121	
Db	61	IQSENMAGPNNASARGFVNVVDIKTALENTCPCGVYSCDIALALASVASYLTGGFSWYL	120	
QY	122	LGRDSTQAFIDLANANLPAFFFTLLPOLKDSFRVNGLNKSSDVALSGGHTFGKNQCRPI	181	
Db	121	LGRDSTLANLAGANSAIPSPFEGLSNITSKPSAVGLN-TNDVLVALSGAHTFGRCRGVF	179	
QY	182	MDRLXNENSTGLPPTLTWTTLYLTQIRGCLPLNGNLNLSALVDEPLKRPITFDNKYYNLEQ	241	
Db	180	NNRLEFNGSTGMPPTLNTSLTSLSLQDLCPONGSASTITNDLSTPDADFNNYFANLQSN	239	
QY	242	KGLIISDDELSSPATPTIPLVRSEFANSTQTFENAFVPEANDRMGNTIPLGTGQIRLN	301	
Db	240	NGLIISDDELSTLIGSA-TIAVYTSFASNOITLFFQFAQSMINMGNISPLTGSNGEIRLD	298	
QY	302	CRVYNSN	308	
Db	299	CKVYDGS	305	
RESULT	9			
PERX_TOBAC	PERX_TOBAC	STANDARD	PRT	324 AA.
AC	P11965			
DT	01-OCT-1989 (Rel. 12, Created)			
DT	01-OCT-1989 (Rel. 12, Last sequence update)			
DT	01-APR-1990 (Rel. 14, Last annotation update)			
DE	LIGNIN FORMING ANIONIC PEROXIDASE PRECURSOR (BC 1.11.1.7).			
OS	Nicotiana tabacum (Common tobacco).			
OC	Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: Asteridae: euasterids I: Solanales;			
OC	Solanaceae: Nicotiana.			
OC	[1]			
RN	SEQUENCE FROM N.A.			
RP	Lagrimini L.M., Moyer M., Rothstein S.;			
RA	"Molecular cloning of complementary DNA encoding the lignin-forming			
RT	peroxidase from tobacco: molecular analysis and tissue-specific			
RT	expression."			
RT	Proc. Natl. Acad. Sci. U.S.A. 84:7542-7546(1987).			
CC	-1 FUNCTION: THIS ENZYME PLAYS AN INTEGRAL ROLE IN SECONDARY CELL			
CC	WALL BIOSYNTHESIS BY THE POLYMERIZATION OF CINNAMYL ALCOHOLS			
CC	INTO LIGNIN AND BY FORMING RIGID CROSS-LINKS BETWEEN CELLULOSE,			
CC	PECTIN, HYDROXY-PROLINE-RICH GLYCOPROTEINS, AND LIGNIN.			
CC	-1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.			
CC	-1- COFACTOR: BINDS HEME.			

CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: J02979; AAA34108.1; -  
DR PIR: A39889; A39889.  
DR HSP: P00433; 1ATU.  
DR PRAM: PF00141; peroxidase; 1.  
DR PRINTS: PR00458; PEROXIDASE.  
DR PRINTS: PR00461; PLPEROXIDASE.  
DR PROSITE: PS00435; PEROXIDASE\_1; 1.  
DR PROSITE: PS00436; PEROXIDASE\_2; 1.  
KW Oxidoreductase; Peroxidase; Heme; Glycoprotein; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 324 LIGNIN FORMING ANIONIC PEROXIDASE.  
FT MOD\_RES 23 23 PYRROLIDONE CARBOXYLIC ACID.  
FT ACT\_SITE 60 60 BY SIMILARITY.  
FT ACT\_SITE 64 64 DISTAL HISTIDINE (BY SIMILARITY).  
FT ACT\_SITE 189 189 PROXIMAL HISTIDINE (HEME AXIAL LIGAND)  
FT (BY SIMILARITY).  
FT DISULFID 33 111 BY SIMILARITY.  
FT DISULFID 66 71 BY SIMILARITY.  
FT DISULFID 117 320 BY SIMILARITY.  
FT DISULFID 196 228 BY SIMILARITY.  
FT CARBOHYD 35 35 POTENTIAL.  
FT CARBOHYD 150 150 POTENTIAL.  
FT CARBOHYD 207 207 POTENTIAL.  
SQ SEQUENCE 324 AA; 34674 MW; 0F1F03927F47A669 CRC64;  
  
Query Match 48.9%; Score 786.5; DB 1; Length 324;  
Best Local Similarity 51.0%; Pred. No. 5.1e-59;  
Matches 156; Conservative 56; Mismatches 89; Indels 5; Gaps 5;  
  
OY 2 QLTPTFYDSCPNVSNIVROTIVNELSDPRIASILRLHDFCVNGCDASILLDNITS 61  
DB 23 QLSATPFYDTTCPNVTSIVRGVMDORORDARAGAKIIRLHPHDFVNGCDSILLD-TDG 81  
  
OY 62 FRTKEDAFGNANSARGFVIDRMKAAYESACPTVSCADLLITTAAGSVTLAAGPSWRP 121  
DB 82 TOTKRDAPANY-GAGGEDIYDIDTALENVCPGVASCADILALSELGVYLAKEPSWVL 140  
  
OY 122 LGRDLSLOAFILANANLPAPFTLPLKDSFRNVGLNRSSDLVALSGHTFGKNCRFI 181  
DB 141 FGRDLSLANSRGSNDIPSPFETLAVIPIQFTKGM-DLTDVALSGAHFGRAGCTF 199  
  
OY 182 MDRLYNSNGLPPTLTNTTYLQTLRGICPLNGUL-SALVDFDLRTPIFNKYYVLEE 240  
DB 200 EQRLEFNFGSCNPDLTYDAFLQTLGICPOGGNNGNFTMLDISTPNDYFTNIQS 259  
  
OY 241 OKGLIOSDQELFSSPDATDITPLVRSFANSQTFEFNAFVEAMGNTPTLTGOGORL 300  
DB 260 NQGLQTDQDELFT-SSGATIAVNRVAGSOTOFFDFVSMKLGINSPLTGINGQIRT 318  
  
OY 301 NCRVYN 306  
DB 319 DCKRVN 324

GN PNC1.  
OS *Arachis hypogaea* (Peanut).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae;  
OC Papilionoideae; Arachis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91062381.  
RA Buffard D., Breda C., van Huystee R.B., Asemota O., Pierre M.,  
RA Dang Ha D.B., Esnault R.;  
RT "Molecular cloning of complementary DNAs encoding two cationic  
RT peroxidases from cultivated peanut cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:8874-8878(1990).  
RN [2]  
RP REVISION TO 47.  
RA Esnault R.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RX X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
RP MEDLINE; 96398617.  
RA Schuller D.J., Ban N., van Huystee R.B., McPherson A., Poulos T.L.;  
RT "The crystal structure of peanut peroxidase.";  
RL Structure 4:311-321(1996).  
CC -1- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,  
CC BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD  
CC WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE  
CC DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.  
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) -> OXIDIZED DONOR + 2 H(2)O.  
CC -1- COFACTOR: BINDS HEME.  
CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES (CLASS III).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M37636; AAB06183.1; -  
DR PIR: A38265; A38265.  
DR PDB: 1SCH; 11-JUL-96.  
DR PRAM: PF00141; peroxidase; 1.  
DR PRINTS: PR00458; PEROXIDASE.  
DR PRINTS: PR00461; PLPEROXIDASE.  
DR PROSITE: PS00435; PEROXIDASE\_1; 1.  
DR PROSITE: PS00436; PEROXIDASE\_2; 1.  
KW Oxidoreductase; Glycoprotein; Peroxidase; Heme; Multigene family;  
KW Signal; 3D-structure.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 316 CATIONIC PEROXIDASE 1.  
FT MOD\_RES 23 23 PYRROLIDONE CARBOXYLIC ACID.  
FT ACT\_SITE 60 60 BY SIMILARITY.  
FT ACT\_SITE 64 64 DISTAL HISTIDINE (BY SIMILARITY).  
FT ACT\_SITE 191 191 PROXIMAL HISTIDINE (HEME AXIAL LIGAND).  
FT CARBOHYD 82 82 POTENTIAL.  
FT CARBOHYD 166 166 POTENTIAL.  
FT CARBOHYD 207 207 POTENTIAL.  
FT DISULFID 33 113 POTENTIAL.  
FT DISULFID 66 71  
FT DISULFID 119 312  
FT DISULFID 198 223  
SQ SEQUENCE 316 AA; 33517 MW; 2CC271F8B8BC9F0 CRC64;

Query Match 46.4%; Score 745.5; DB 1; Length 316;  
Best Local Similarity 49.2%; Pred. No. 1.4e-55;  
Matches 150; Conservative 45; Mismatches 99; Indels 11; Gaps 3;  
  
OY 2 QLTPTFYDSCPNVSNIVROTIVNELSDPRIASILRLHDFCVNGCDASILLDNITS 61  
DB 23 QLSATPFYDTTCPNVTSIVRGVMDORORDARAGAKIIRLHPHDFVNGCDSILLDTSN 82

QY 62 FTRKDAFGNANSARGFPVIDRMKAVERACPTVSCADLLTTAAGOSVTLAGPSMRVP 121  
 DB 83 FTGEKTAGPNANSIRGEVITDIKQSVESICPGVYSCADILAAVARSVALGASMNVL 142  
 QY 122 LGRDSTLQAFDLANANLPAPEFTLPOLKDSFRVNGLNRRSDVALSGHTEGNCQRFI 181  
 DB 143 LGRDSTLQAFDLANANLPAPEFTLPOLKDSFRVNGLNRRSDVALSGHTEGNCQRFI 201  
 QY 182 MDRLYNFSNTGLPDPPTLNTTYLTQTLRGICPLNGNLSALVDFDLRTPTIFDNKYYVNLREQ 241  
 DB 202 RTRIYNSN-----IDPTTAKSLQANCPSVGGDTLNSPFDVYTPPKFDMAYINLRNK 254  
 QY 242 KGLIQSDQELFSSPDADTDTIPLVRSFANSTQTFENAFVEMDRNGNTPLTGTGQIRLN 301  
 DB 255 KGLIHSDDQLF-----NGVSTDSQVTAYSNNMATEFNTDFGNAMIKMGNLSPLTGTSGQIRTN 311  
 QY 302 CRVYN 306  
 DB 312 CRKTN 316

RESULT 11

PER2\_CUCSA STANDARD; PRT; 292 AA.  
 ID PER2\_CUCSA

AC P19135;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-MAR-1992 (Rel. 21, Last annotation update)  
 DE PEROXIDASE 2 (EC 1.11.1.7) (FRAGMENT).  
 OS Cucumis sativus (Cucumber).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Cucurbitales;  
 OC Cucurbitaceae; Cucumis.  
 RN [1]

SEQUENCE FROM N.A.

RX MEDLINE: 91346662.

RA Morgens P.H., Callahan A.M., Dunn L.J., Abeles F.B.;

RT Isolation and sequencing of cDNA clones encoding ethylene-induced

RL Plant Mol. Biol. 14:715-725(1990).

CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.

CC -1- COFACTOR: BINDS HEME.

CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: M32742; AAA33121.1; -.

CC PIR: S11870; S11870.

CC HSSP: P00433; IATY.

CC PFAM: PF00141; peroxidase; 1.

CC PROSITE: PS00435; PEROXIDASE\_1; 1.

CC PROSITE: PS00436; PEROXIDASE\_2; 1.

CC Oxidoreductase; Peroxidase; Heme; Glycoprotein.

KW NON\_TER 1

FT ACT\_SITE 36 36 BY SIMILARITY.

FT ACT\_SITE 38 38 DISTAL HISTIDINE (BY SIMILARITY).

FT ACT\_SITE 164 164 PROXIMAL HISTIDINE (HEME AXIAL LIGAND).

FT CARBOHYD 68 68 POTENTIAL.

FT CARBOHYD 139 139 POTENTIAL.

FT CARBOHYD 179 179 POTENTIAL.

SQ SEQUENCE 292 AA; 31856 MM; 98FD89EF01CE6888 CRC64;

Query Match 45.9%; Score 737.5; DR 1; Length 292;  
 Best Local Similarity 52.8%; Pred. No. 5; 8e-55;  
 Matches 159; Conservative 34; Mismatches 99; Indels 9; Gaps 6;

QY 6 TFYDNCSPVNSIVRDTIYNELRSDPRIAASILRLHFDCEVNGCDASILLDNTPSRTF 65  
 DB 1 TFYDNCSPVNSIVRDTIYNELRSDPRIAASILRLHFDCEVNGCDASILLDNTPSRTF 60  
 QY 66 KDAFGNANSARGFPVIDRMKAVERACPTVSCADLLTTAAGOSVTLAGPSMRVPYIGRR 125  
 DB 61 LAAGGNAN-ITGFEIVANNIKAAVERACPTVSCADILAIASVGVNLAGPCWCVQIGRR 119  
 QY 126 DSIQAFDLNANLPAPEFTLPOLKDSFRVNGLNRRSDVALSGHTEGNCQRFIDRL 185  
 DB 120 DSRANLQAGIDGLPSFEVNTQLRKFRDVLDD-STDLVALSGAHTFGKRCQF-FDRR 177  
 QY 186 YNFSNTGLPDPPTLNTTYLTQTLRGICPLNGNLSALVDFDLRTPTIFDNKYYVNLREQGLI 245  
 DB 178 LANSN-----PDSTLNPRYAQQLRQAC--SSGRDFEVNLDPTTPPKFDMAYINLRNK 232  
 QY 246 QSDQELFSSPDADTDTIPLVRSFANSTQTFENAFVEMDRNGNTPLTGTGQIRLNCRVY 305  
 DB 233 TSDQVLSHTP-GEDTVKIVNLFASQNOFFESFGOSMIMNGNIQPILOGNGEIRSNCRRL 291  
 QY 306 N 306  
 DB 292 N 292

RESULT 12

PERL\_WHEAT STANDARD; PRT; 312 AA.  
 ID PERL\_WHEAT

AC Q05855;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 DE PEROXIDASE PRECURSOR (EC 1.11.1.7).  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.  
 RN [1]

SEQUENCE FROM N.A.

RC STRAIN-CV. CHEYENNE;

RX MEDLINE: 91363838.

RA Hertig C., Rehmann G., Bull J., Mauch F., Dudler R.;

RT "Sequence and tissue-specific expression of a putative peroxidase

RL Plant Mol. Biol. 16:171-174(1991).

CC -1- FUNCTION: INVOLVED IN DEFENSE RESPONSE TO POWDERY MELDREW FUNGUS.

CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.

CC -1- COFACTOR: BINDS HEME.

CC -1- SUBCELLULAR LOCATION: SECRETED (PROBABLE).

CC -1- TISSUE SPECIFICITY: ROOT.

CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: X53675; CAA37713.1; -.

CC PIR: S13325; S13325.

CC HSSP: P22195; ISCH.

CC PFAM: PF00141; peroxidase; 1.

CC PRINTS: PR00456; PEROXIDASE.

CC PROSITE: PS00435; PEROXIDASE\_1; 1.

CC PROSITE: PS00436; PEROXIDASE\_2; 1.

CC Oxidoreductase; Peroxidase; Heme; Glycoprotein; Multigene family;

KW Signal 1

FT CHAIN 24 23 POTENTIAL.

FT ACT\_SITE 61 61 PEROXIDASE.

FT ACT\_SITE 65 65 BY SIMILARITY.

FT ACT\_SITE 65 65 DISTAL HISTIDINE (BY SIMILARITY).

```

FT   ACT_SITE      185      185      PROXIMAL HISTIDINE (HEME AXIAL LIGAND)
FT   DISULFID      34       107      (BY SIMILARITY).
FT   DISULFID      67       70       BY SIMILARITY.
FT   DISULFID      113      307      BY SIMILARITY.
FT   DISULFID      192      218      BY SIMILARITY.
FT   CARBOHYD      262      262      POTENTIAL.
SQ   SEQUENCE      312 AA; 32381 MW; FF468A33F3DC68F7 CRC64;

Query Match      45.4%; Score 729.5; DB 1; Length 312;
Best Local Similarity 49.7%; Pred. No. 3e-54;
Matches 152; Conservative 41; Mismatches 96; Indels 17; Gaps 5;

QY   2 QLTPEFYDNCSPNWSNIVRDTIYNELNSDPRIASILRLHFDCEVNCDCASILDNTTTS 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   24 QLSSTFYDTCSPRALVAIKSGVAAVSSDPRMGASILRLHFHDFC--GCDASVLLTG--- 78

QY   62 FRTEKDEAGNMSNRAGFVIDIRKRAAVESACPRTYSCADLTITAAQOSVTLAAGPSMKVP 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   79 --MEQNAPPNVSGSLRGFEVIDINIKQLLESVCVKQTVSCADLTITVAARDVVALGSPSWTP 136

QY   122 LGRRLSLOAFIDLANANLPAFPFTLPOLKDSFRRNLGRSSDLVALSGCHTFGKKOGRFI 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   137 LGRDSTTASIASLANSDLPGPSSSSRSQSEAAFLKKNLN--TYDMAVLSAHTIGKACCSNF 195

QY   182 MDRLYNFSNTGLPPTLTNTTYTLQTLRGCLPLNGNLNALVDFDLRPTTFDNKKYVNLREQ 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   196 RTRIRG-----GDPIINTAFATSLKANCPOSGGNTNLANDTMTPNMFDNAYTYNTLLSQ 249

QY   242 KGLIISDOLEFSSPATDTITPLVNSFANSSTOTFFNAFVANDRMGNITPLTGTGQIRLN 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   250 KGLIHSDDVLEFNNETTDMT---VRNFASNAAFSSAFITAMIKMGINAIPLTGTGQIRLS 306

QY   302 CRYVNS 307
    ||| |||
Db   307 CSKYNS 312

RESULT 13
PER2_ORYSA
ID   PER2_ORYSA      STANDARD:      PRT:      314 AA.
AC   P37835;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   PEROXIDASE PRECURSOR (EC 1.11.1.7).
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
    Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPONBARE; TISSUE=ROOT;
RA   Hori M., Sasaki T., Minobe Y.;
RL   Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
CC   -1- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,
    BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD
    WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE
    DEPENDENT ON EACH ISOTYPE/ISOPORN IN EACH PLANT TISSUE.
CC   -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.
CC   -1- COFACTOR: BINDS HEME.
CC   -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@sib-sib.ch).
CC
EMBL; D16442; BAA03911.1; -.
HSSP; P22195; 1SCH.
OR

```

DR PFAM: PF00141; peroxidase; 1.  
DR PRINTS: PR00459; PEROXIDASE.  
DR PRINTS: PR00461; PLPEROXIDASE.  
DR PROSITE: PS00435; PEROXIDASE\_1; 1.  
DR PROSITE: PS00436; PEROXIDASE\_2; 1.  
KW oxidoreductase; glycoprotein; Peroxidase; Heme; Multigene family;  
KW Signal.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 314 PEROXIDASE.  
FT MOD\_RES 24 24 PYROLIDONE CARBOXYLIC ACID (POTENTIAL).  
FT ACT\_SITE 61 61 BY SIMILARITY.  
FT ACT\_SITE 65 65 DISTAL HISTIDINE (BY SIMILARITY).  
FT ACT\_SITE 187 187 PROXIMAL HISTIDINE (HEME AXIAL LIGAND).  
FT CARBOHYD 148 148 POTENTIAL.  
FT CARBOHYD 169 169 POTENTIAL.  
FT CARBOHYD 203 203 POTENTIAL.  
FT CARBOHYD 274 274 POTENTIAL.  
FT CARBOHYD 309 309 POTENTIAL.  
SQ SEQUENCE 314 AA; 32606 MW; 232c874730dc7fb CMC64;

Query Match 44.4%; Score 714; DB 1; Length 314;  
Best Local Similarity 48.2%; Pred. No. 6e-53;  
Matches 144; Conservative 47; Mismatches 94; Indels 18; Gaps 5.

QY 2 QLPETFDNSCPNVSNIYDTIVNELRSDPRIAASILRLHFHDFCVNGCDASILLDNQTS 61  
||: |||| |||| : : : : ||: ||: |||| |||| ||||: ||  
Db QLSATFEDTSCPNALSTIKSAVTAAVNSEPRMGASLVRHFHDFCVNGCDASVLLSG--- 80

QY 62 FRTEKDAFGNANSARGFVIDIMKRAAVSACPRYSCADLLTAAQSYTLAGGFSWRKP 121  
||: |||| |||| : : : : ||: ||: |||| |||| ||||: ||  
Db --QEQMAGPNAGSLRGFVVDINIKQYVAISQYVSCADILLAAVARDVVALGGSPWTVL 138

QY 122 LGRDLSLOAFDILNANLPAPFETLPQKDSFRVNGLRSSDLVALSGGTFGKNOCRPT 181  
||||| : : : |||| : : ||: ||: |||| |||| ||||: ||  
Db LGRDSTTANNSQANTDLPAPSSSLAEIIGNFSRGLD-VYDMVALSGAHITIGQAQQNF 197

QY 182 MDRLNFSNTGLPPTLTLYLQTLRGLC--PLNGNLSALVDFDLRPTIIFDNKYVME 239  
||||| : : : |||| : : ||: ||: |||| |||| ||||: ||  
Db RDRLYNETN-----IDSEFATALKANCPRPTGSGDNLAPLDPTTPNAPFSAYYTNL 250

QY 240 EQKGLIGSDQELFSSPDATDTIPLVRSFANSTQTGFNAFVEMADMGNTPLTGTGGQIR 299  
||||: |||| : : : |||| : : ||||: |||| |||| ||||: ||  
Db 251 SNKGKIGSDQVLEFNGSGTDNT---VRNFSSTNAFNSAFYTAAMVKMGNISPLTGTGGQIR 307

QY 300 LNCRAVYN 306  
||| ||  
Db 308 LNCSTKVN 314

RESULT 14  
PERL\_HORVU STANDARD; PRT; 315 AA.  
AC P27337;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE PEROXIDASE 1 PRECURSOR (EC 1.11.1.7).  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. PALLAS / P-01; TISSUE=SEEDLING LEAF.  
RA Thorold-Christensen H., Brandt J., Cho B.H., Rasmussen S.K.,  
RA Giegersten P.L., Smedegaard-Petersen V., Collinge D.B.;  
RT "cDNA cloning and characterization of two barley peroxidase  
transcripts induced differentially by the powdery mildew fungus  
Erysiphe graminis".  
RL Physiol. Mol. Plant Pathol. 40:395-409(1992).  
CC -1: FUNCTION: INVOLVED IN DEFENSE RESPONSE TO POWDERY MILDEW FUNGUS.  
CC -1: CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.  
CC -1: COFACTOR: BINDS HEME.

CC Solanaceae; Nicotiana.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV, XANTHI NC;  
RX MEDLINE; 93041285.  
RA Criqui M.-C., Plesse B., Durr A., Marbach J., Parmentier Y.,  
RT Jamet E., Fleck J.;  
RT "Characterization of genes expressed in mesophyll protoplasts of  
RT Nicotiana sylvestris before the re-initiation of the DNA  
RT replicational activity.";  
RL Mech. Dev. 38:121-132(1992).  
CC -1- FUNCTION: THIS ENZYME PLAYS AN INTEGRAL ROLE IN SECONDARY CELL  
CC WALL BIOSYNTHESIS BY THE POLYMERIZATION OF CINNAMYL ALCOHOLS,  
CC INTO LIGNIN AND BY FORMING RIGID CROSS-LINKS BETWEEN CELLULOSE,  
CC PECTIN, HYDROXY-PROLINE-RICH GLYCOPROTEINS, AND LIGNIN.  
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.  
CC -1- COFACTOR: BINDS HEME.  
CC -1- TISSUE SPECIFICITY: MESOPHYLL PROTOPLASTS AND TO A MUCH LESSER  
CC EXTENT, ROOTS AND GERMINATING SEEDS.  
CC -1- DEVELOPMENTAL STAGE: BEFORE RE-INITIATION OF THE DNA REPLICATIONAL  
CC ACTIVITY.  
CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC

DR	PFWM:	PF00141;	peroxidase_1.
DR	PRINTS:	PR00456;	PEROXIDASE.
DR	PRINTS:	PR00461;	PLPEROXIDASE.
DR	PROSITE:	PS00435;	PEROXIDASE_1; 1.
DR	PROSITE:	PS00436;	PEROXIDASE_2; 1.
KW	Oxidoreductase:	Peroxidase;	Heme; Glycoprotein; Signal.
FT	SIGNAL	1	27
FT	CHAIN	28	322
FT	MOD_RES	28	28
FT	ACT_SITE	65	65
FT	ACT_SITE	69	69
FT	ACT_SITE	196	196
FT	DISULFID	38	118
FT	DISULFID	71	76
FT	DISULFID	124	318
FT	DISULFID	203	229
FT	CARBOHYD	213	213
SO	SEQUENCE	322 AA;	34646 MW; OC154FCDC2596449 CRC64;

Query Match                  43.7%;    Score 702;   DB 1;   Length 322;  
Best Local Similarity      48.9%;    Pred. No. 6.3e-52;  
Matches 149; Conservative     42; Mismatches 104; Indels    10; Gaps        5;

QY	2 QLTPEFYVNSCPNVSNIYRDITIVNELRSDPRIASILRLHHPDFVGNGCDASILDDNTTS	61
DB	28 QLSTAEFDNTCPNALMTIRISVKQAISERMMASLRLHFHDCEFGCCDSILLDETPSS	87
QY	62 FRTKEKDAFGNMSARGFEVIDIMRKAAVESACPPIRYSCADLLTIIAAGSVTLAGSPSMRV	121
DB	88 IESKETALPNIGSGARGREIIDEAKREVBEIKICPGVYSCHDILTVAARDASAAGVPSPMYK	147
QY	122 LGRDSDLGAFDLIANANLPAPFTLLPOLKDSFRNVGLNRSSDYVALSGHTFGKNOCRFI	181
DB	148 LGRRDSLTAASKTIAETLDGPFPEDPLNLRISSFASKGLS -TFDMVALSAHAHTIGAQCFLF	206
QY	182 MDRLNFNSMGCLPDPITLVTYQLTRGLCPLINGNISALVDLFLLRPPTIFDKKYVNLEEQ	241

Db	207	RDRY--SN---CTDIDAGFASTRRRROCPQEGENGNIAPLDLVTNPQFDNNYFKNLIOK	260
QY	242	KGLIQSDQELFSSPDATDTIPIVRSFANSTOTFENAFVEAMDRMGNTTPLTGTGQIRLN	301
Db	261	KGLIQSDQVLENG-GSTDNI--VSEISNSARAFSSDPAAAMIKMGDISPLSGONGIRKY	317
QY	302	CRVYN	306
Db	318	CGSYN	322

Search completed: October 4, 2000, 13:04:57  
Job time: 1688 sec

**THIS PAGE BLANK (USPTO)**





```
Db 31 QLRPTFYDNCSPVNTIVRETYNELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 90
OY 62 FRTKEADAFGNANSARGPVIDIRMKAAVESACPRVSCADLLTTIAAOSVTLAGPSMRVP 121
Db 91 FRTKEADAFGNANSARGPVIDIRMKAAVESACPRVSCADLLTTIAAOSVTLAGPSMRVP 150
OY 122 LGRDLSIQAFALANLANLPAPFFTLPLQKDSFRVNGINRSSDLVALSGHTFGKNCQCFI 181
Db 151 LGRDLSIQAFALANLANLPAPFFTLPLQKDSFRVNGINRSSDLVALSGHTFGKNCQCFI 210
OY 182 MDRLYNFSNTGLPDPPTLNTTYLOTLRGLCPUNGNLSALVDFDLRTPTIPFNKYYVNLLEQ 241
Db 211 LDRFYNSNTGLPDPPTLNTTYLOTLRGLCPUNGNRSALVDFDLRTPTIPVFNKYYVNLLEK 270
OY 242 KGLIQSDQELFSSPDATDITPLVRSFANSQTQTFENAFVEMADMGNITPLTGTGGIIRLN 301
Db 271 KGLIQSDQELFSSPDATDITPLVRSFANSQTQTFENAFVEMADMGNITPLTGTGGIIRLN 330
OY 302 CRVYNSNS 309
Db 331 CRVYNSNS 338

RESULT 2
O43732 PRELIMINARY; PRT; 352 AA.
AC Q43732;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, last annotation update)
DE PEROXIDASE PRECURSOR (EC 1.11.1.7) (MTELOPEROXIDASE).
GN YPR9 OR PRX3 OR PRX3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-VARIOUS STRAINS;
RA CAPELLI N., TOGNOLLI M., FLACH J., OVERNEY S., PENEL C., GREPPIN H.,
RA SIMON P.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN RP SEQUENCE FROM N.A.
RC STRAIN-VARIOUS STRAINS;
RA WESTERGAARD L., ABELSKOV A.K., JENSEN R.B., HANSEN L.N.,
RA RASMUSSEN S.K.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RN RP SEQUENCE FROM N.A.
RA TOGNOLLI M., GREPPIN H., SIMON P.;
RT peroxidase.
RT "Structure of the gene encoding Arabidopsis thaliana Prx3
peroxidase."
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.
CC -1- COFACTOR: HEME.
DR EMBL: X98315; CAA66959.1; -
DR EMBL: X98777; CAA67313.1; -
DR EMBL: A113036; CAB37193.1; -
DR HSSP: P00433; ZATU.
DR MENDEL: 6566; Arabid. Ypr9.6566.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PROSITE: PS00436; PEROXIDASE_2; 1.
DR PFAM: PF00141; peroxidase: 1.
DR PRINTS: PR00458; PEROXIDASE.
DR PRINTS: PR00461; PEROXIDASE.
DR SIGNAL: 1 29 POTENTIAL.
FT CHAIN 30 337 POTENTIAL.
SQ SEQUENCE 352 AA; 38873 MW; 19DBBFDI CRC32;
```

```
Query Match 89.2%; Score 1434; DB 10; Length 352;
Best Local Similarity 88.6%; Pred. No. 2.5e-121;
Matches 273; Conservative 16; Mismatches 19; Indels 0; Gaps 0;

OY 2 QLRPTFYDNCSPVNSIVRDTYNELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 61
Db 30 QLRPTFYDNCSPVNTIVRETYNELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 89
OY 62 FRTKEADAFGNANSARGPVIDIRMKAAVESACPRVSCADLLTTIAAOSVTLAGPSMRVP 121
Db 90 FRTKEADAFGNANSARGPVIDIRMKAAVESACPRVSCADLLTTIAAOSVTLAGPSMRVP 149
OY 122 LGRDLSIQAFALANLANLPAPFFTLPLQKDSFRVNGINRSSDLVALSGHTFGKNCQCFI 181
Db 150 LGRDLSIQAFALANLANLPAPFFTLPLQKDSFRVNGINRSSDLVALSGHTFGKNCQCFI 209
OY 182 MDRLYNFSNTGLPDPPTLNTTYLOTLRGLCPUNGNLSALVDFDLRTPTIPFNKYYVNLLEQ 241
Db 210 MDRLYNFSNTGLPDPPTLNTTYLOTLRGLCPUNGNQVTLVDFDLRTPTIPFNKYYVNLLEK 269
OY 242 KGLIQSDQELFSSPDATDITPLVRSFANSQTQTFENAFVEMADMGNITPLTGTGGIIRLN 301
Db 270 KGLIQSDQELFSSPDATDITPLVRSFANSQTQTFENAFVEMADMGNITPLTGTGGIIRLN 329
OY 302 CRVYNSNS 309
Db 330 CRVYNSNS 337

RESULT 3
O80912 PRELIMINARY; PRT; 349 AA.
ID O80912;
AC O80912;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, last annotation update)
DE PEROXIDASE.
GN YPR9 OR T19C21.12.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-CV, COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RL "Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC004683; AAC28765.1; -
DR HSSP: P00433; IATU.
DR MENDEL: 31568; Arabid. Ypr9.31568.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PROSITE: PS00436; PEROXIDASE_2; 1.
DR PFAM: PF00141; peroxidase: 1.
DR PRINTS: PR00458; PEROXIDASE.
DR PRINTS: PR00461; PEROXIDASE.
DR PEROXIDASE.
SQ SEQUENCE 349 AA; 38100 MW; 8103AA13 CRC32;

Query Match 68.9%; Score 1108.5; DB 10; Length 349;
Best Local Similarity 70.5%; Pred. No. 4.9e-92;
Matches 215; Conservative 31; Mismatches 58; Indels 1; Gaps 1;

OY 2 QLRPTFYDNCSPVNSIVRDTYNELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 61
Db 30 QLRPDTFYRCPPIFNIGDTYNELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 89
OY 62 FRTKEADAFGNANSARGPVIDIRMKAAVESACPRVSCADLLTTIAAOSVTLAGPSMRVP 121
```

ID	Accession	Species	Length	Score	Similarity	Conservative	Mismatches	Indels	Gaps
Db	90	FRTEDDAAPNKNKSVSGFDYIDNMKAATIERACPRIVSCADITITIASOISVLLSGGFMMPVP	149	67.1%	Score 1079.5; DB 10;	68.2%; Pred. No. 2e-89;	59;	1;	1.
Qy	122	IGRRDLSAFLDILAAANLPAPEFTLLPOLKDSFRNGLNNSDVLVLSGGHTFGKMQCRFI	181						
Db	150	IGRRDSVEAFFLAANTALPSPRESTLTQLKTAADVGLNRPDVLVLSGGHTFGKMQCQCV	209						
Qy	182	MDRLYNFSNTGCLPDPPTLNTTYLTQTLRGCLPLNGNLISALVDEDPDLRPTTIFDNKYYVNLBEO	241						
Db	210	TPRLYNFNGTNPDSPLNPTTYLVELEIRRLCPONGNGTVALNFDPSVMPTEFDROXYTNLLNG	269						
Qy	242	KGLIOSDDELSPDPTPIPLVRSFANSTQFFNFAFPAEMRMGNITPLTGTQOIRLN	301						
Db	270	KGLIOSDDELSPDPTPIPLVNOYSNTTFVGFAPADAMIRMGNLRLKPLTGTQOIRLN	328						
Qy	302	CRVYN 306							
Db	329	CRVYN 333							
RESULT	4								
ID	080913	PRELIMINARY:	PRT:	349 AA.					
AC	080913:								
DT	01-NOV-1998	(TREMBLrel. 08, Created)							
DT	01-NOV-1998	(TREMBLrel. 08, Last sequence update)							
DT	01-NOV-1998	(TREMBLrel. 12, Last annotation update)							
DE	PEROXIDASE.								
GN	YPR9 OR T19C21.13.								
OS	Arabiidopsis thaliana (Mouse-ear cress).								
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;								
OC	euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;								
OC	core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;								
OC	Arabiidopsids.								
RA	SEQUENCE FROM N.A.								
RC	STRAIN=CV. COLUMBIA;								
RA	ROUNDSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,								
RA	STYES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,								
RA	SOBERVILLE C.R., VENTER J.C.;								
RT	"Arabiidopsis thaliana chromosome II BAC T19C21 genomic sequence.";								
RL	Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.								
DR	EMBL; AC004683; AAC28766.1; -.								
DR	HSSP; P00433; IATJ.								
DR	MENDEL; J1569; Arabid.Ypr9.J1569.								
DR	PROSITE; PS00435; PEROXIDASE_1; 1.								
DR	PROSITE; PS00436; PEROXIDASE_2; 1.								
DR	PFAM; PF00141; PEROXIDASE; 1.								
DR	PRINTS; PR00458; PEROXIDASE.								
DR	PRINTS; PR00461; PLPEROXIDASE.								
DR	PEROXIDASE.								
SO	SEQUENCE 349 AA: 38108 MW: 0792EBF6 CRC32:								
Query Match	67.1%;	Score 1079.5;	DB 10;	Length 349;					
Best Local Similarity	68.2%;	Pred. No. 2e-89;							
Matches 208;	Conservative 37;	Mismatches 59;	Indels 1;	Gaps 1.					

Query	242	KGLOSDELFSSPATDTIPILVBSFANSTQFFNAFVEMDRMGNTIPITGTQOIRLN	301
Db	270	KGLOSDELFSTPGA-DTIPLVNYSDDMSVFFRAFIADIMIRGNLRITGTQOGEIRON	328
Qy	302	CRVYN 306	
Db	329	CRVYN 333	
RESULT	5		
Q43100			
AC	043100	PRELIMINARY:	PRT: 343 AA.
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation update)	
DE	PEROXIDASE (EC 1.11.1.7) (MYELOPEROXIDASE).		
GN	YPR9.		
OS	Populus trichocarpa (Western balsam poplar).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;		
OC	core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=XYLEM;		
RA	CHRISTENSEN J.H., BAUW G., BOERJAN W., VAN MONTAGU M.;		
RL	Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.		
CC	-I- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.		
CC	-I- COFACTOR: HEME.		
DR	EMBL: X97349; CAA66035.1; -		
DR	HSSP: P00433; ZATJ.		
DR	MSDEL: 10232; PopTr:Ypr9.10232.		
DR	PROSITE: PS00435; PEROXIDASE_1; 1.		
DR	PROSITE: PS00436; PEROXIDASE_2; 1.		
DR	PFAM: PF00141; peroxidase; 1.		
DR	PRINTS: PR00458; PEROXIDASE.		
DR	PRINTS: PR00461; PEROXIDASE.		
KM	peroxidase; oxidoreductase.		
SQ	SEQUENCE 343 AA; 36568 MW; BFECE2DB CRC32;		
Query Match	65.4%;	Score 1052;	DB 107; Length 343;
Best Local Similarity	64.5%;	Pred. NO. 5.8e-87;	
Matches 198;	Conservative 43;	Mismatches 66;	Indels 0; Gaps 0;
Qy	2	QUTPTFYDNCSPVNSINIVADITVNELRSDPRITAAIILRHFFHDFCVNGCDASILDNNTS	61
Db	25	QUTPTFYDNCSPVNSINIVADITVNELRSDPRITAAIILRHFFHDFCVNGCDASILDNNTS	84
Qy	62	FTEEDATGNNANSANGFVIDRMAKAVESACRITYSCADLITIAAQSYTLAGPSMRVP	121
Db	85	IYSEKEAGNNNSANGFEVVDTKMALLKESACPAIVSCADILITIAEESVVLGGPNMTVP	144
Qy	122	LGRROSLAFLDLNANALPAPFETLIPOLKDSRVNGLNNSDVLVALSGHTGKNGCRPT	181
Db	145	LGRROSTARSRAANAALPAPFETLIPOLKDSRVNGLNNSDVLVALSGHTGKNGCRPT	204
Qy	182	MDRLVNSFTGLPDPDLNTTYLQTLRGCLPLNGNLSAIVDFDLRTPTTIPDNKYVYNLEEO	241
Db	205	MDRLVNSFTGLPDPDLNTTYLQTLRGCLPLNGNLSAIVDFDLRTPTTIPDNKYVYNLEEO	264
Qy	242	KLLISDDELSSPATDTIPILVBSFANSTQFFNAFVEMDRMGNTIPITGTQOIRLN	301
Db	265	KLLISDDELSSPATDTIPILVBSFANSTQFFNAFVEMDRMGNTIPITGTQOIRLN	324
Qy	302	CRVYN 308	
Db	325	CRVYN 331	
RESULT	6		
Q40950			

```
ID 040950 PRELIMINARY; PRT; 343 AA.
AC 040950;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PEROXIDASE.
GN YPR9.
OS Populus nigra (lombardy poplar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
RN [1]
RP SEQUENCE FROM N.A.
RA SAKUMA Y., AZUMA T., KATO Y., KOJIMA Y., MIURA K.;
RL "Poplar peroxidase gene responding to cytokinin treatment.";
EMBL: D83225; BA11853.1; -.
DR HSSP: P00433; ZATU.
DR MENDEL: 12193; Popul; Ypr9; 12194.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PROSITE: PS00436; PEROXIDASE_2; 1.
DR PFAM: PF00141; peroxidase; 1.
DR PRINTS: PR00458; PEROXIDASE.
DR PRINTS: PR00461; PLPEROXIDASE.
KW Peroxidase.
SQ SEQUENCE 343 AA; 36532 MW; 1645A1A5 CRC32;

Query Match 65.0%; Score 1046; DB 10; Length 343;
Best Local Similarity 63.2%; Pred. No. 2e-86;
Matches 194; Conservative 47; Mismatches 66; Indels 0; Gaps 0;

QY 2 QLTPTFYDNCSPVSNIVRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 61
   |||||:||||:| | | |||||:|||||:|||||:|||||:|||||:
Db 25 QLTPTFYDQTCPNVSSIIIRVITETIVSDPRIASILRLHHDCLVNGCDOSILLDMTDT 84
   |||||:||||:| | | |||||:|||||:|||||:|||||:|||||:
QY 62 FRTFKDAFGNANSARGPVYIDRMKAAYESACPRIVSCADLITTAAGSVTLAGSPSRVP 121
   :||:| | |||||:| | | | | | | | | | | | | | | | | | | |
Db 85 IESKEAGGNNNSARGPEVYIDRMKALLESACPAVSCADLITTAABESVTLAGSPNMTVP 144
   :||:| | |||||:| | | | | | | | | | | | | | | | | | | |
QY 122 LGRDSTLQAFDLANANLPAFFTLPLQKDSFRVNGLRSSDVALSGHFGKNOGRFI 181
   ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db 145 LGRDSTTASDAANANLPAFFTLPLQKDSFRVNGLRSSDVALSGAHFGKACSTP 204
   ||||| | | | | | | | | | | | | | | | | | | | | | | |
QY 182 MDRLYNFSNTGAPDPPTLNTTYLTQTLRGICPLNGNISALVDFDLTPTIFDNKYVNEEQ 241
   ||:|:| | | | | | | | | | | | | | | | | | | | | | | |
Db 205 DFRLEFDNMSGAPDPSPDPTLTLAQLGELCPGCGKRSYITDDLTTPAFDSNYTSNLQGN 264
   ||:|:| | | | | | | | | | | | | | | | | | | | | | | |
QY 242 KGLIQSQDELSSPDADTDTPLVNSFANSYQTFEFNAVEAMDRNGNITPLTGGQIRLN 301
   :||:| | | | | | | | | | | | | | | | | | | | | | | |
Db 265 RGLLQTOGELFSTPGADVDIALVNAFSAQTAFFESFESMIRMGNLSPLTGTGEIRLN 324
   :||:| | | | | | | | | | | | | | | | | | | | | | | |
QY 302 CRVYNSN 308
   |||||:|
Db 325 CRVYNNAN 331
   |||||:|

RESULT 7
Q40949 PRELIMINARY; PRT; 343 AA.
ID 040949;
AC 040949;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PEROXIDASE.
GN YPR9.
OS Populus nigra (lombardy poplar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
RN [1]
RP SEQUENCE FROM N.A.
RA SAKUMA Y., AZUMA T., KATO Y., KOJIMA Y., MIURA K.;
```

```
RT "Poplar peroxidase gene responding to cytokinin treatment.";
RL Submitted (JAN-1996) to the EMBL/GenBank/DDJ databases.
DR EMBL: D83224; BA11852.1; -.
DR HSSP: P00433; ZATU.
DR MENDEL: 12193; Popul; Ypr9; 12193.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PROSITE: PS00436; PEROXIDASE_2; 1.
DR PFAM: PF00141; peroxidase; 1.
DR PRINTS: PR00458; PEROXIDASE.
DR PRINTS: PR00461; PLPEROXIDASE.
KW Peroxidase.
SQ SEQUENCE 343 AA; 36529 MW; E1AB207D CRC32;

Query Match 64.9%; Score 1044; DB 10; Length 343;
Best Local Similarity 63.8%; Pred. No. 3e-86;
Matches 196; Conservative 44; Mismatches 67; Indels 0; Gaps 0;

QY 2 QLTPTFYDNCSPVSNIVRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 61
   |||||:||||:| | | |||||:|||||:|||||:|||||:|||||:
Db 25 QLTPTFYDQTCPNVSSIIIRVITETIVSDPRIASILRLHHDCLVNGCDOSILLDMTDT 84
   |||||:||||:| | | |||||:|||||:|||||:|||||:|||||:
QY 62 FRTFKDAFGNANSARGPVYIDRMKAAYESACPRIVSCADLITTAAGSVTLAGSPSRVP 121
   :||:| | |||||:| | | | | | | | | | | | | | | | | | | |
Db 85 IYSEKEAGGNNNSARGPEVYIDRMKALLESACPAVSCADLITTAABESVTLAGSPNMTVP 144
   :||:| | |||||:| | | | | | | | | | | | | | | | | | | |
QY 122 LGRDSTLQAFDLANANLPAFFTLPLQKDSFRVNGLRSSDVALSGHFGKNOGRFI 181
   ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db 145 LGRDSTTASDAANANLPAFFTLPLQKDSFRVNGLRSSDVALSGAHFGKACSTP 204
   ||||| | | | | | | | | | | | | | | | | | | | | | | |
QY 182 MDRLYNFSNTGAPDPPTLNTTYLTQTLRGICPLNGNISALVDFDLTPTIFDNKYVNEEQ 241
   ||:|:| | | | | | | | | | | | | | | | | | | | | | | |
Db 205 DFRLEFDNMSGAPDPSPDPTLTLAQLGELCPGCGKRSYITDDLTTPAFDSNYTSNLQGN 264
   ||:|:~| | | | | | | | | | | | | | | | | | | | | | | |
QY 242 KGLIQSQDELSSPDADTDTPLVNSFANSYQTFEFNAVEAMDRNGNITPLTGGQIRLN 301
   :||:| | | | | | | | | | | | | | | | | | | | | | | |
Db 265 QGLLQTOGELFSTPGADVDIALVNAFSAQTAFFESFESMIRMGNLSPLTGTGEIRLN 324
   :||:| | | | | | | | | | | | | | | | | | | | | | | |
QY 302 CRVYNSN 308
   |||||:|
Db 325 CRVYNNAN 331
   |||||:|

RESULT 8
Q43101 PRELIMINARY; PRT; 343 AA.
ID 043101;
AC 043101;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PEROXIDASE (EC 1.11.1.7) (MNELOPEROXIDASE).
GN YPR9.
OS Populus trichocarpa (Western balsam poplar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
RN [1]
RP SEQUENCE FROM N.A.
RA CHRISTENSEN J.H., BAUW G., BOERJAN W., VAN MONTAGU M.;
RL Submitted (APR-1996) to the EMBL/GenBank/DDJ databases.
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) -> OXIDIZED DONOR + 2 H(2)O.
DR EMBL: X87350; CA66036.1; -.
DR HSSP: P00433; ZATU.
DR MENDEL: 10233; Popul; Ypr9; 10233.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PROSITE: PS00436; PEROXIDASE_2; 1.
DR PFAM: PF00141; peroxidase; 1.
DR PRINTS: PR00458; PEROXIDASE.
DR PRINTS: PR00461; PLPEROXIDASE.
KW Peroxidase; Oxidoreductase.
SQ SEQUENCE 343 AA; 36750 MW; 35EB6C5 CRC32;
```

[illegible]

Dd	85	I ESEKEAGNNNSARGFEVVDVRMKALLLESACPAIVSCADILTTIAAESVYLAAGPNNHTVP	144
Oy	122	LGRDLSLOAFDLANANI.PAPFFETLLPOLKDSFRVNGLRSSDLYALSGGTFEGNKQCFI	181
		:    :    :    :    :    :    :    :	
Dd	145	LGRRSSTJASRAAANAASIPAPFLTLTDOLRESFTFVGNLNNTDLYALSGAHFGCAKCSF	204
Oy	182	MDRLLNFENMTGLPDDTLWTMYTLOTLRGCPLNGNLSALVDPDDLRTPTFDKKYYVNLEEQ	241
		:    :    :    :    :    :    :    :	
Dd	205	NFRILYDFENGTAPODTLPDLPELAAQLCECPGGNDSTYTDLDTTPPAOFDSNYYSNLCCN	264
Oy	242	KGLISDDELFSPPATPTITIEVRFSANSTOTFEFAFVEADMRCMNITPLETGTOGOTIRN	301
		:    :    :    :    :    :    :    :    :	
Dd	265	RGLTOTDELFESTPADDIVIALYNFASNOTAFFESPVESHMRGMNLSPLTGTGEIRLN	324
Oy	302	CRVANSN 308	
		:	
Dd	325	CSVVAN 331	
RESULT	10		
ID	Q43050	PRELIMINARY;	PRT; 343 AA.
AC	Q43050:		
DT	01-NOV-1996	(TREMBUREL_01, Created)	
DT	01-NOV-1996	(TREMBUREL_01, last sequence update)	
DT	01-NOV-1999	(TREMBUREL_12, last annotation update)	
DE	PEROXIDASE [EC 1.11.1.7].		
GN	YP99 OR PRXAAA.		
OS	Populus kitchakamensis (Aspen).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;		
OC	coryle dicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 95375232.		
RA	OSAKABE K., KOYAMA H., KAWAI S., KATAYAMA Y., MOROHOSHII N.;		
RT	"Molecular cloning of two tandemly arranged peroxidase genes from		
RT	Populus kitchakamensis and their differential regulation in the stem."		
RL	Plant Mol. Biol. 28:677-689(1995).		
DR	EMBL: D38051; BAA07241.1; -		
DR	HSSP: P00433; ZATJ.		
DR	MENDEL: 12185; Popk1.Ypr9.12185.		
DR	PROSITE: PS00435; PEROXIDASE_1; 1.		
DR	PROSITE: PS00436; PEROXIDASE_2; 1.		
DR	PFAM: PF00141; Peroxidase; 1.		
DR	PRINTS: PR00458; PEROXIDASE.		
DR	PRINTS: PR00461; PLPEROXIDASE.		
KM	Peroxi-dase; Oxidoreductase.		
FT	CHAIN 25 343		
SO	SEQUENCE 343 AA; 37220 MW; 83FC34D1 CRC32;		
Query Match	62.3%;	Score 1002;	DB 10; Length 343;
Best Local Similarity	62.2%;	Pred. NO. 1.8e-82;	
Matches 191;	Conservative 38;	Mismatches 78;	Indels 0; Gaps 0;
Oy	2	OLTFFYVNSCGNVNINVRDIIVNELRSDPRIAISILRLHFHDCFNVCDCASILLDNPTS	61
		:    :    :    :    :    :    :    :	
Dd	25	OLTFFYDETCTCYVSIIRGVIAETLIRDPRIGASILRLHHDCTVNCDCSILLDKAT	84
Oy	62	FTEERDAFCMANASARGFPVIDRMKAAVBSACPDRVSCADLLTTIAAQSVTLAGPSWRVP	121
		:    :    :    :    :    :    :    :	
Dd	85	IDTEKEALANNNSARGFOVDIMKERLEGVCPCATVSCADILIAAESEVSVLAGEPMPIR	144
Oy	122	LGRRSLOAFDLANANI.PAPFFETLLPOLKDSFRVNGLRSSDLYALSGGHFGKNQCFI	181
		:    :    :    :    :    :    :    :	
Dd	145	LGRRSSTJANRTAAAFETPGODTLIERLRSRTVVGLNNTDLYALSGAHFGCAKCSF	204
Oy	182	MDRLLNFENMTGLPDDTLWTMYTLOTLRGCPLNGNLSALVDPDDLRTPTFDKKYYVNLEEQ	241
		:    :    :    :    :    :    :    :    :	
Dd	205	IDRLYNFNMTGLPDDTLDTTYLATIQRICFCPOGKGCTVALADDPTTPDGDFNMNYSNLQAS	264
Oy	242	KGLISDDELFSPPATPTITIEVRFSANSTOTFEFAFVEADMRCMNITPLETGTOGOTIRN	301

```
Db 265 KGLQSDQELFSTPEADITELVDISTDETAFFESFVSMIRGNLSPLTGTGETRLN 324
QY 302 CRYVNSN 308
Db 325 CRAVNAD 331

RESULT 11
ID 043051 PRELIMINARY: PRT: 314 AA.
AC 043051:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PEROXIDASE (EC 1.11.1.7) (MYELOPEROXIDASE) (FRAGMENT).
GN YPR9.
OS Populus kitakamiensis (Aspen).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=STEM.
RA OSAKABE K., KOYAMA H., KAWAI S., KATAYAMA Y., MOROHOSHI N.;
RL Plant Sci. 103:167-175(1994).
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.
CC -1- COFACTOR: HEME.
DR EMBL: D30652; BAA0634.1; -.
DR HSSP: P00433; 2ATJ.
DR MENDEL: 12186; POPKI.YPR9:12186.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PFAM: PF00141; PEROXIDASE_2; 1.
DR PEROXIDASE; Oxidoreductase.
KW NON_TER
FT SEQUENCE 314 AA; 33471 MW; 9501F77A CRC32;
SQ

Query Match 61.9%; Score 996; DB 10; Length 314;
Best Local Similarity 62.6%; Pred. No. 5.6e-82;
Matches 189; Conservative 41; Mismatches 72; Indels 0; Gaps 0;

QY 7 FYDNCSPVNSIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLDNSTFETEK 66
Db 1 FYDTCPCNVSTIIRVDYETETFLASDPRIASILRLHFHDCFVNGCDGSLLDNSDTIYSEK 60

QY 67 DARGNANSARGFVYIDRMKAIVESACPRVSCADLLTIAAQSVTLAAGSWMKQCFRMDRL 126
Db 61 EAGGNNSARGFVYIDRMKAIVESACPRVSCADLLTIAEESVLAAGSNMVPPLGRD 120

QY 127 SLQAFDLANANPAPFETLPOLKDSFRNGLRRSSDVLVALSGHFGKNGCFRMDRL 186
Db 121 STASRDAAANAPLPAPRITIDOLRESTVNGLNNSDVLVALSGAHTFGRAKCSFTPRRL 180

QY 187 NESNTGLPDPPLNTVYLOTLRGLCPLNGNLSALVDFDLRTPTIEDNKYYVLEEQKLIQ 246
Db 181 DFNSTGAPDQSLDPTLLAQLQELCPQGGNSVLTLDLDTTPDAFDSNYSNLQNGQLIQ 240

QY 247 SDDELSSPADPTITPLVRSFANSTOTFFNAFVPEAMDRMGNITPLVGTQOQIRLNCRVN 306
Db 241 TDDVLTSTPCADYIALVNAFSAHQTAFFESFASMSIRMGNLRLPLGTGTGEIRLNCRVN 300

QY 307 SN 308
Db 301 AN 302

RESULT 12
ID 043102 PRELIMINARY: PRT: 345 AA.
AC 043102:
DT 01-NOV-1996 (Tremblrel. 01, Created)
```

```
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PEROXIDASE (EC 1.11.1.7) (MYELOPEROXIDASE).
GN YPR9.
OS Populus trichocarpa (Western balsam poplar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=XYLEM.
RA CHRISTENSEN J.H., BAUM G., BOERJAN W., VAN MONTAGU M.;
RL Submitted (APR-1996) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.
CC -1- COFACTOR: HEME.
DR EMBL: X97351; CAA66037.1; -.
DR HSSP: P00433; 1ATJ.
DR MENDEL: 10234; POPKI.YPR9:10234.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PROSITE: PS00436; PEROXIDASE_2; 1.
DR PFAM: PF00141; PEROXIDASE_1.
DR PRINTS: PR00461; PEROXIDASE.
DR PEROXIDASE; Oxidoreductase.
KW SEQUENCE 345 AA; 36648 MW; AFA48C40 CRC32;
SQ

Query Match 61.9%; Score 995.5; DB 10; Length 345;
Best Local Similarity 62.0%; Pred. No. 7.1e-82;
Matches 191; Conservative 43; Mismatches 73; Indels 1; Gaps 1;

QY 2 QLPRTFYDNCSPVNSIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLDNSTTS 61
Db 29 QLPRTFYDTCPCNVSTIIRGVLAQALQTPDRIASILRLHFHDCFPDGGCGSLLDNSTD 88

QY 62 FRTFKDAFGNANSARGFPYIDRMKAIVESACPRVSCADLLTIAAQSVTLAAGSPKMRP 121
Db 89 IESEKEAPNNNSARGFVYIDRMKAIVESACPRVSCADLLTIAEESVLAAGSPMTVP 148

QY 122 LGRDLSQAFDLANANPAPFETLPOLKDSFRNGLRRSSDVLVALSGHFGKNGCFRMDRL 181
Db 149 LGRDLSITANRSCANSSTIPASSELAVLWASKFAVAGLNTSSDVLVALSGAHTFGRAOCLNF 208

QY 182 MDRLNFSNTGLPDPPLNTVYLOTLRGLCPLNGNLSALVDFDLRTPTIEDNKYYVLEEQ 241
Db 209 ISRLYNFSGSGNDPPLNTVYLAQLQELCPQGGNRSVLTLDLDTTDTFGNANFSNLQTN 268

QY 242 KGLQSDQELFSSPDAITITPLVRSFANSTOTFFNAFVPEAMDRMGNITPLVGTQOQIRL 301
Db 269 EGLQSDQELFSTTGA-DTIAIVNFSNOTAFFESFVSMIRGNLSPLTGTGETRLN 327

QY 302 CRYVNSN 309
Db 328 CRIVNST 335

RESULT 13
ID 043049 PRELIMINARY: PRT: 347 AA.
AC 043049:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PEROXIDASE PRECURSOR (EC 1.11.1.7) (PEROXIDASE) (MYELOPEROXIDASE).
GN YPR9 OR PRXA3A.
OS Populus kitakamiensis (Aspen).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 95375232.
RA OSAKABE K., KOYAMA H., KAWAI S., KATAYAMA Y., MOROHOSHI N.;
```

"Molecular cloning of two tandemly arranged peroxidase genes from *Populus kirkamienensis* and their differential regulation in the stem.";  
 RL Plant Mol. Biol. 28:677-689(1995).

RP SEQUENCE OF 40-347 FROM N.A.  
 RC TISSUE-STEM;  
 RA OSAKABE K., KOYAMA H., KAWAI S., KATAYAMA Y., MOROHOSHI N.;  
 RL Plant Sci. 103:167-175(1994).  
 CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.  
 DR -1- COFACTOR: HEME.  
 CC  
 DR EMBL: D38050; BAA07240.1; -;  
 DR HSSP: P00433; ZATU.  
 DR MENDEL: 12184; Popk1.Ypr9;12184.  
 DR PFAM: PF00141; peroxidase; 1.  
 DR PRINTS: PR00458; PEROXIDASE.  
 DR PRINTS: PR00461; PLPEROXIDASE.  
 KW Oxidoreductase; Peroxidase.  
 FT CHAIN 30 347 POTENTIAL.  
 SQ SEQUENCE 347 AA; 37079 MW; 30D3FE8B CRC32;

Query Match 60.7%; Score 975.5; DB 10; Length 347;  
 Best local Similarity 62.0%; Pred. No. 4.5e-80;  
 Matches 191; Conservative 39; Mismatches 77; Indels 1; Gaps 1;

OY 2 QLTPTFYDNCSPVNSIVRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDTTTS 61  
 Db 30 QLSPTFYDNCSPVNSIVRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDTTDT 89  
 OY 62 FRTKDAFGNANSARGFVIDRKAAYESACPTVSCADLLITAAQSVTLAGSPSWRV 121  
 Db 90 ISEKEAEPNNNSVGRGDVDDMKALENACPGIVSCADILATIAEBSVCLAGSPSWTV 149  
 OY 122 LGRDLSIAFLDANANLPAFFTLPOLKDSFRVNGLNSSDLVALSGGTFGKNCRT 181  
 Db 150 LGRDLSIAFRSAGNSALBSPFASLVLSKFAVGLDTSDDLVALSGAHTFGAACSS 209  
 OY 182 MDRLYNSNTGLDPPPTNTYTLTGLRGLCPNLGNLSALVDFDLRTPTIFPNKYVNEQ 241  
 Db 210 NLRLYFSGSGNDPPLNTTYTLTGLRGLCPNLGNLSALVDFDLRTPTIFPGNTFSLQTN 269  
 OY 242 KGLIQDQELFSSPDATDITPLVRSFANSTOTFEFNAFEAMDGMNTPLTGTGOTIRLN 301  
 Db 270 EGLRSDQELFSTGA-DTIDIYNNFSSNOTAFEFESVVMIRGNISPLTGTGDEIRLN 328  
 OY 302 CRVYNSNS 309  
 Db 329 CRRVNDNS 336

## RESULT 14

ID 042905 PRELIMINARY; PRT; 359 AA.  
 AC 042905;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)  
 DE PEROXIDASE, PRECURSOR (EC 1.11.1.7) (MYELOPEROXIDASE).  
 GN YPR9 OR FLXPER1.  
 OS Linum usitatissimum (Flax) (Linseed).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids I; Malpighiales; Linaceae; Linum.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-STORMONT CIRRUS; TISSUE=LEAF, AND STEM;  
 RA OMANN F., TYSON H.;  
 RL Aust. J. Plant Physiol. 23:773-789(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-STORMONT CIRRUS; TISSUE=LEAF, AND STEM;  
 RA OMANN F., TYSON H.;

RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.  
 DR -1- COFACTOR: HEME.  
 CC  
 DR EMBL: L07554; AABA7602.1; -;  
 DR HSSP: P00433; ZATU.  
 DR MENDEL: 11888; Linus.Ypr9;11888.  
 DR PFAM: PF00141; peroxidase; 1.  
 DR PRINTS: PR00458; PEROXIDASE.  
 DR PRINTS: PR00461; PLPEROXIDASE.  
 KW Signal; Peroxidase; Oxidoreductase.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 28 359 PEROXIDASE.  
 SQ SEQUENCE 359 AA; 38197 MW; F8479674 CRC32;

Query Match 59.4%; Score 955.5; DB 10; Length 359;  
 Best local Similarity 59.1%; Pred. No. 3e-78;  
 Matches 182; Conservative 45; Mismatches 80; Indels 1; Gaps 1;

OY 2 QLTPTFYDNCSPVNSIVRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDTTTS 61  
 Db 28 QLTPTFYDNCSPVNSIVRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDSAT 87  
 OY 62 FRTKDAFGNANSARGFVIDRKAAYESACPTVSCADLLITAAQSVTLAGSPSWRV 121  
 Db 88 ILSEKQALGNNSVGRGDVDDMKQVEACPGVSCADILATIAEBSVCLAGSPSWAV 147  
 OY 122 LGRDLSIAFLDANANLPAFFTLPOLKDSFRVNGLNSSDLVALSGGTFGKNCRT 181  
 Db 148 LGRDLSIAFRSAGNSALBSPFASLVLSKFAVGLDTSDDLVALSGAHTFGAACSS 207  
 OY 182 MDRLYNSNTGLDPPPTNTYTLTGLRGLCPNLGNLSALVDFDLRTPTIFPNKYVNEQ 241  
 Db 208 VGRLYNFSNTGCPDPTNTTYTLTGLRGLCPNLGNLSALVDFDLRTPTIFPNKYVNEQ 267  
 OY 242 KGLIQDQELFSSPDATDITPLVRSFANSTOTFEFNAFEAMDGMNTPLTGTGOTIRLN 301  
 Db 268 EGLRSDQELFSTP-GSDTIELVRFANOTAFQSFVSMIRGNIPPPGSPSEIRN 326  
 OY 302 CRVYNSNS 309  
 Db 327 CRRVNSNS 334

## RESULT 15

ID 09XFL3 PRELIMINARY; PRT; 341 AA.  
 AC 09XFL3;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)  
 DE PEROXIDASE 1 PRECURSOR (FRAGMENT).  
 GN FBPI.  
 OS Phaseolus vulgaris (Kidney bean) (French bean).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. MONCALM.  
 RA BLEE K.A., JUNE S.C., RICHARD G., BOWWELL G.P.;  
 RT "Molecular cloning and differential expression of six members of the  
 peroxidase family in French bean."  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF149277; AAD37427.1; -;  
 KW Peroxidase.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 341 AA; 36626 MW; C147F3A0 CRC32;

Query Match 57.5%; Score 924; DB 10; Length 341;  
 Best local Similarity 59.4%; Pred. No. 1.9e-75;

	Matches 183;	Conservative 44;	Mismatches 79;	Indels 2;	Gaps 2;
QY	2	QLTPTFEYDNSCPVNSIVIRDPYIVNELRSDPRIASILRLHFHDCFYNGCDASILLDNTTS	61		
Db	16	QLDPSPFYRNTGCPVSHSVREVIRNVSKSDPRMLASLIRLHFHDCFYNGCDASILLDNTDT	75		
QY	62	FRTKDAFGNANSARGFPVIDRMAAVESACPRVSCADLLTIAAQSVTLAGGPPSWRP	121		
Db	76	IVSEQALPPIINSIRGLADVYVNIKTAVENACPGVSCADILLTIAAFISVYLAQGPDKVP	135		
QY	122	LGRDLSLOAEFLDANANLPAPFTFLPOLKDSFRNVGLNRSSDIYALSGHTEGKNOCRT	181		
Db	136	LGRKDSLTAARLTANONLPAPFFNLTLKAAFAVQGLN-TTDLVALSGAHTFGAQCSTP	194		
QY	182	MDRLYNFSNTGLPDPPTLNTTYTLQTLRGCLPLNGNLSALVDFDLRTPTIFDNKYYVNL	241		
Db	195	VNRLYNFSNTGNPDPTLNTTYTLQTLRAVCPNGGGTINLTNFDPTTPDKFDKNYNSNL	254		
QY	242	KGLIQSDQELFSSPDATDTIPLVRSFANSTQTFEFAFEVAMDRMGNTTPLTGTOGIRLN	301		
Db	255	KGLIQSDQELFSTIGA-DFTIDIVNRFSSNOTLFEFESKAMIMKGNIGVLTGSQGEIRKQ	313		
QY	302	CRVYVNSNS	309		
Db	314	CNEFVNGNS	321		

Search completed: October 4, 2000, 12:39:40  
Job time: 21393 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:36:40 ; Search time 73.09 Seconds  
(without alignments)  
350.531 Million cell updates/sec

Title: US-09-246-451-11

Perfect score: 2180  
Sequence: 1 TTTETIOSNANLAPLPHPVPE.....IVSGVQALPLVWDPAFTRKAV 414

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2176	99.8	415	1	O4PSCP
2	377	17.3	393	1	C70929
3	372	17.1	410	2	S39924
4	366	16.8	396	1	B69851
5	364	16.7	397	1	S51594
6	360	16.5	398	1	H70807
7	346.5	15.9	404	2	T30231
8	340	15.6	403	2	B35401
9	338.5	15.5	428	1	F70729
10	335.5	15.4	410	1	O4B56M
11	333.5	15.3	405	1	B42606
12	333	15.3	410	1	E69611
13	330	15.1	406	1	A35401
14	329	15.1	411	1	JC5859
15	326	15.0	406	1	A48495
16	324.5	14.9	376	1	G69679
17	319.5	14.7	405	1	H70752
18	318	14.6	310	2	T44767
19	318	14.6	428	1	A42971
20	316.5	14.5	385	1	G69594
21	316.5	14.5	399	1	A53578
22	316	14.5	412	2	S24750
23	313.5	14.4	410	2	JC4287
24	304	13.9	438	1	E70515
25	303	13.9	402	2	JC5151
26	303	13.9	404	1	JC5150
27	303	13.9	417	1	S49051
28	296.5	13.6	408	2	PD0007
29	296.5	13.6	408	2	JC5674

30	294.5	13.5	337	2	T30601	cytochrome P450 hy
31	293.5	13.5	406	1	S18531	cytochrome P450 er
32	291.5	13.4	433	1	B70677	cytochrome P450 Rv
33	287	13.2	381	1	S15809	cytochrome P450 Cy
34	277.5	12.7	310	2	T44857	probable hydroxyla
35	266	12.2	412	1	B40634	erythromycin monoo
36	265	12.2	414	1	E70708	cytochrome P450 Rv
37	262.5	12.0	411	2	T36526	probable cytochrom
38	261.5	12.0	386	2	T30235	cytochrome P450 -
39	255	11.7	406	3	T17487	cytochrome P450 hy
40	253	11.6	401	1	T40208	cytochrome P450 Bu
41	250	11.5	411	1	JC4003	cytochrome P450 -
42	239	11.0	433	2	T44587	cytochrome P450 ho
43	236	10.8	402	1	A70707	cytochrome P450 Rv
44	234	10.7	396	1	H70730	cytochrome P450 Rv
45	226	10.4	391	3	T17486	cytochrome P450 hy

## ALIGNMENTS

## RESULT 1

O4PSCP  
camphor 5-monooxygenase (EC 1.14.15.1) cytochrome P450 101 - Pseudomonas putida plasm  
N:Alternate names: cytochrome P450-CAM  
C:Species: Pseudomonas putida  
C>Date: 30-Apr-1982 #sequence,revision 31-Dec-1993 #text\_change 03-Mar-2000  
C:Accession: A25660; S34614; C60886; A00194  
R:Unger, B.P.; Gunsalus, I.C.; Silgar, S.G.  
J. Biol. Chem. 261, 1158-1163, 1986  
A:Title: Nucleotide sequence of the Pseudomonas putida cytochrome P-450-cam gene and  
A:Reference number: A94678; MUID:6111751  
A:Accession: A25660  
A:Molecule type: DNA  
A:Residues: 1-415 <UNG>  
A:Cross-references: GB:M12546; MID:g151114; PIDN:AAA25760.1; PID:g151115  
R:Atamaki, H.; Koga, H.; Sagara, Y.; Hosoi, M.; Horiuchi, T.  
Biochim. Biophys. Acta 1174, 91-94, 1993  
A:Title: Complete nucleotide sequence of the 5'-exo-hydroxycamphor dehydrogenase gene  
A:Reference number: S34613; MUID:93326643  
A:Accession: S34614  
A>Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-42 <ARA>  
A:Experimental source: PpG1, ATCC 17453; CAM plasmid  
R:Romeo, C.; Moriwaki, N.; Yasunobu, K.T.; Gunsalus, I.C.; Koga, H.  
J. Protein Chem. 6, 253-261, 1987  
A:Title: Identification of the coding region for the putidaredoxin reductase gene fro  
A:Reference number: A60886  
A:Accession: C60886  
A:Molecule type: DNA  
A:Residues: 408-415 <ROM>  
R:Haniu, M.; Armes, L.G.; Yasunobu, K.T.; Shastri, B.A.; Gunsalus, I.C.  
J. Biol. Chem. 257, 12664-12671, 1982  
A:Title: Amino acid sequence of the Pseudomonas putida cytochrome P-450. II. Cyanogen  
A:Reference number: A00194; MUID:33030788  
A:Accession: A00194  
A:Molecule type: protein  
A:Residues: 2-55, 58-276, 'Q', 278-361, 'S', 363-407, 'N', 409-415 <HAN>  
C:Genetics:  
A:Gene: camC; CYP101  
A:Genome: plasmid  
C:Function:  
A:Description: catalyzes hydroxylation of camphor to yield 5-exo-hydroxycamphor; elec  
C:Superfamily: Pseudomonas plasmid camphor 5-monooxygenase; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygena  
F:246-380/domain: cytochrome P450 homology <CYP>  
F:358/binding site: heme iron (Cys) (axial ligand) #status experimental

Query Match 99.8%; Score 2176; DB 1; Length 415;  
Best Local Similarity 99.8%; Pred. No. 5.1e-159;  
Matches 413; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY	1	TTETIOSNANLAPLEPHPEHLEVFEDMYNPSNISAGVOEMAYLOESNVYDIWTRONG	60
Db	2	TTETIOSNANLAPLEPHPEHLEVFEDMYNPSNISAGVOEMAYLOESNVYDIWTRONG	61
OY	61	GHMIATRGOLIREAEDYRHFSSSECPFIPIREAGEAYDIPIASMPPPEOROFALANQVYG	120
Db	62	GHMIATRGOLIREAEDYRHFSSSECPFIPIREAGEAYDIPIASMPPPEOROFALANQVYG	121
OY	121	MPVYDKLENRIOELACSLIESLRPOGOCNFTEDYAEPPRIKIFMLLAGLPREDIPIHLKYL	180
Db	122	MPVYDKLENRIOELACSLIESLRPOGOCNFTEDYAEPPRIKIFMLLAGLPREDIPIHLKYL	181
OY	181	TDOMTRPBGSMTEFAKKEALYDYLPIITEORROROKGTALISIVANGOVNGRPISDEAKR	240
Db	182	TDOMTRPBGSMTEFAKKEALYDYLPIITEORROROKGTALISIVANGOVNGRPISDEAKR	241
OY	241	MCGILLVAGLDTPVVFLEFSMEFLAKSPENHOELERPERIPACECELLRFRFSIYAODRI	300
Db	242	MCGILLVAGLDTPVVFLEFSMEFLAKSPENHOELERPERIPACECELLRFRFSIYAODRI	301
OY	301	LTSDEYEFHGVOLKKODQILLPOMLSGLDERKNACPMHVDFSROKAVSHTFGHGSHLQIG	360
Db	302	LTSDEYEFHGVOLKKODQILLPOMLSGLDERKNACPMHVDFSROKAVSHTFGHGSHLQIG	361
OY	361	HLAREIIIVTLEKWIETRIPIDEFISIAPGAOIOHKSIGVSVQVALPLVMDPATTKAV	414
Db	362	HLAREIIIVTLEKWIETRIPIDEFISIAPGAOIOHKSIGVSVQVALPLVMDPATTKAV	415

RESULT 2  
C70929  
cytochrome P450 RV1785c - Mycobacterium tuberculosis (strain H37Rv)  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Mycobacterium tuberculosis  
C:Date: 10-Sep-1999 #sequence.revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: C70929  
R:cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squires, S.  
Native 393, 537-544, 1998  
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: C70929  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-393 <COL>  
A:Cross-references: GB:AL022021; GB:AL123456; NID:93250699; PIDN:CA17707.1; PID:el25459  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV1785c  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase

	Query March	17.3%:	Score 377:	DB 1:	Length 393;
	Best Local Similarity	28.7%:	Pred. No. 2.2e-21:		
	Matches 110;	Conservative 70;	Mismatches 183;	Indels 20;	Gaps 10;
OY	31	PSNLACVQEAAMAVLQESNPDLVTWTRCNGSHWTATRGQLIREAYEDYRHFSSSECPFIPR	90		
	:	:	:   :	:   :	:   :
Db	23	PMAADRGV--GWKTLRLDAG--PVPVF--MNGWYL-TREDELTALIRMPKVFSSRAKLQP-	74		
OY	91	EAGEAYDEIPTSMPPPEOROFRALANQVGVPVDKIENRFOELACSGSTIESLRGGCNF	150		
	:	:     :	:   :     :	:   :     :	:   :     :
Db	75	-PGNPLPVPLAFDPPEHTRYRILQPFYSPALSKALPSLRRTVTAMDIALARGCEA	133		
OY	151	TEDVAEPPIPIFMILLGLPEEDIPLHLKYLLDOM----TRDGSGMTFAEAKELXYLIIP	206		
	:	:   :             :	:     :   :             :	:     :   :             :	:     :   :             :
Db	134	MADLANLEFPOLFTVLGLPLEDRRLGMKDVAIAMSDRHPTEADVAARRELLEYLTA	193		
OY	207	IIEQRROKPGTDAISIVANGCVNGRPITSDEAKRMCGILLVGLDTVVNFLSFSMEFLAK	266		

```

Db      194  MVAEERRRNPQDVLVSQ---QIGEDPSEIEVLGTLSHLLIAGIDVTAAVGFSLLEIAR 250
QY      267  SPENHQLIEPPEPIPAACEELLR-RESYIADGKILTSYDFEFGYQKKEDQILLQMLS 325
Db      251  RPQLNMYLRNDPKQIRVFETIEIVLEPSPAPAPVPTTEPVTGGMTLPASSPRLICMAAV 310
QY      326  GIDERNKNAQPMHNVDFSRQKVSHTTFEGGSHLGLQHLIAREIIVTLTKEMGTRIPDSIAP 385
Db      311  NRDSGDAMSTDELVMDCGVNHNMGFGGGRPHRCGLSHIARLETLVLVGEIMNQIPDELAP 370
QY      386  --GAQIQKSGSIVSGQALPLVM 406
Db      371  DYAPETIRPSSSFA-LKMLPLRW 392

```

RESULT 3  
S39924  
cytochrome P450meg - Bacillus megaterium (ATCC 13368)  
C:Species: Bacillus megaterium  
A:Variety: ATCC 13368  
C:Date: 07-Oct-1994 #sequence-revision 23-Feb-1996 #text-change 04-Mar-2000  
C:Accession: S39924; S32216  
R:Rauschenbach, R.; Isernhagen, M.; Noeske-Junghlut, C.; Boldol, W.; Stewart, G.  
Mol. Gen. Genet. 241: 170-176, 1993  
A:Title: Cloning sequencing and expression of the gene for cytochrome P450meg, the st  
A:Reference number: S39924; MUID:94049677  
A:Accession: S39924  
A:Molecule type: DNA  
A:Residues: 1-410 <RAU>  
A:Cross-references: EMBL:Z21972; NID:g288298; PIDN:CAN79985.1; PID:g288300  
A:Experimental source: ATCC 13368  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; heme; iron; metalloprotein  
;35/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	17.1%	Score 372;	DB 21;	Length 410;
Best Local Similarity	30.1%	Pred. NO. 5.5e-21;		
Matches 109; Conservative	53;	Mismatches 180;	Indels 20;	Caps 7

```

OY 63 MIARGOLIREAVEDYRHFSSE-----CEPIREAGEAVDFIP-TSMDDPEORORALA 115
Db 44 MWVRKEYEDVKNVLSDYKHFFSVKRTTITSVGTDBSEESVEKQIOTESDPPDKRRSL 103
OY 116 NOVVGMPEVVDKLENRIOELACLSTLESRPOGOCNFETDVAEPPIRIFMLAGLPEDIP 175
Db 104 AAFPTPSLQWMERIEIDEIDELGQMDGSTEIDIVASLSPRIIYIMADLMKVPKDRY 163
OY 176 HLKLTIDMTPRDGSMTFAE-----AKENALDYULPITEQRKRGTAISIVANGOV 228
Db 164 LFKMWVDTLFLPFDEREKOEVDKLQYAAKEUYQULPIYVOKRLNPADDIISDLKSEY 223
OY 229 NGRRTTSDENAKRMGGLLVGSDLPVNVFSLFSM-EFLAKSPENROELIERPEIIPAAEE 287
Db 224 DGEHPTDDEVVRTTLLIGAVETTSLLNLNFSYSLLYDKEUYOELIENWLDLPQVAEE 283
OY 288 LLR-RESVLADGRILTSVDYENHGVQLKKGOILLPQMLSGDERRKACRYNHWDFSR-QYV 345
Db 284 MLRRFPNLIKIDRIYKEDNDNLGVELKEGGSUYVWMSANMDEMEDPRTLIIHPRNKK 343
OY 346 SHTEFGHSHLCLQHLARREIIVTLKEMWLRIPDESIAFGAIOIHK---SGIVSGVAIL 402
Db 344 KHLNFGNPRHSCGAPLARLRLAEAKIALAPLAKKFKHLEAVRSPQLEBNTLDSANQOTLSL 403
OY 403 PL 404
Db 404 PL 405

```

RESULT 4  
B69851  
cytochrome P450 yj1B - *Bacillus subtilis*

us-09-246-451-11.rpr

```

A:Accession: S51594
A:Molecule type: DNA
A:Residues: 1_397 <TNO>
A:Cross-references: EMBL:D16098; MID:g286050; PIDN:BAA03672.1; PID:d1004187; PID:g30303
C:Genetics:
A:Gene: mycG
A:Start codon: GTG
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: heme; oxidoreductase
F:231-368/Domain: cytochrome P450 homology <CYP>

Query Match      16.7%; Score 364; DB 1; Length 397;
Best Local Similarity 28.5%; Pred. No. 2,28-20;
Matches 109; Conservative 57; Mismatches 168; Indels 38; Gaps 8;

QY   63 WIATRGILREAYEDYRHFSSECPFI-----PREAGAYDFTPTSDPPEROGFRA 113
    | : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   44 WLWTR-----YEDVNAVLDGRFVRGSPMTRDEPRFPMPVKGGLSLMPPEHSRLRR 96

QY   114 LANVGAMPVVDKLENRROELACSLIESLRPQG-CNTDEDYAEPFRIRIMLAGLDEE 172
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   97 LVNAKAFARRAESLRPRAREIAHELVDMAATGAPADLVAMFAARLPKVIVCELGVPSA 156

QY   173 DIPHLKYLTDMTPPDSCM-----TEFAKEALDYDLPIITEQRQAPEGTDALSIV 223
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   157 D-----HDFETMSGAFLSTAETVAEMQEAENQAVAYMGDILDRRKKEPTDLYSAL 209

QY   224 ANGVNGRPITSDAKRMCGLLVGGDITYVNFLSFSGEFLAKSPENHQELIERPERIPA 283
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   210 VQADDQDSISEQELLDAIGLLVAGYSTTWQIADPFYVILLMTBPRRIQLDDRELPLPS 269

```

```
Db      270 AVEELTKMVPRLGVGTAPPRVAVEDYTLRGVTIRAGEPFLASTGAAANRQAOAPPDADRIDV   329  
| | | | | | | | | | : : : : :  
  
OY     341 SRKVSHTTTEGGHSLCLCGHLLARREILIVTLEKMLTRIIPDSIA-PCAQIQHKSG-IVSG    398  
|         | | | | | | | | | | : : : : :  
Db     330 DRFTNQHGFCFHGHVCGLGARLAVELOVALVELLRPLGRIGRIPIPETQLRMSEGMILLRG    389  
| | | | | | | | | | : : : : :  
  
OY     399 VQAALPYW 406  
| | | | | | | | | | : : : : :  
Db     390 PLPLEPVW 397  
| | | | | | | | | | : : : : :  
  
RESULT        6  
H70807  
cytochrome P450 RV3518C - Mycobacterium tuberculosis (strain H37RV)  
N:Contains: oxidoreductase (EC 1.-.-.)  
C:Species: Mycobacterium tuberculosis  
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999  
C:Accession: H70807  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,  
J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Randream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; PMID:98295987  
A:Accession: H70807  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1..398 <COL>  
A:Cross-references: GB:AL022022; GB:AL123456; MID:g3261554; PIDN:CAA17755_1; PID:e125  
C:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV3518c  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase
```

	Query Match	Best Local Similarity
	16.5%	Score 360;
	28.2%; Pred.	DB 1; Length 398; No. 4.de-20;

Matches	107;	Conservative	69;	Mismatches	169;	Indels	34;	Gaps	8;
OY	55	WTRCN-----	GGHMIATRGQLREAYEDYRHSS-----	ECFPIDREAGEAYDFIPT	101				
Db	26	WMRAQPVFPRDRNGIAASTYQAVIDAEROPELFSNAGGIRPDQALP-----	MMI	76					
OY	102	SMDDPEQGFQFALAAQVGMYPVLDLENRIQELACSLIESLRPOGQCFTEYDAEPPIR	161						
Db	77	DMDDPAHLIRRLVLAAGFTRRKRVKDEASIALCPTLLIDACERECDFVRLAALPLMA	136						
OY	162	IFMLIAGL-PEEDIPHLK-----	YLTDOMTRPDGSMTFEAKKALDYLLPIIEGRQ	213					
Db	137	VIGDMLGVRPGRQDMFLKMSDDLVTFLSSHQDEQFQITM-DAAFAAYNDFTRATIAARRA	195						
OY	214	KPGDAIIVANGOVNGRPITSDCAKRMCGLLVGLDVTVNFVLSFMEFLAKSPENHOE	273						
Db	196	PTDDELIVLVSSEVDGGRSLSDDELVMETLLILIGDETFTHHTLSGTEQLLRNDQDL	255						
OY	274	LIERERIPACEELLRRFSLVAD-GRITSDYEPHGVQALKKQDQILPOMLSIGDEKKN	332						
Db	256	LORDPSLLPGALIEELMRTAPYKKNCRVLTADTEHGTALCGEEMMLFESANDEAVF	315						
OY	333	ACPMHVDPSROKVSHTTGGSHSLCLGOHLARREIIVLKEMLTIPDFSIAPGAQI--Q	390						
Db	316	CEPEKFDVQVRNPNSHLARGFTGHFCLGNOQLARLTELSTMTERTLRRLPDULRLVADDSVPL	375						
OY	391	HKSGIVSGVALPLVMDPA	409						
Db	376	RPANFVSGIESMPVFTPS	394						

RESULT 7  
T30231  
cytochrome P450 - Streptomyces hygroscopicus  
C:Species: Streptomyces hygroscopicus  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: J30231  
R:Rapapido, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun  
Gene 169, 9-16, 1996  
A:Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hy  
A:Reference number: Z20782  
A:Accession: T30231  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-404 <APA>  
A:Cross-references: EMBL:X66780; NID:9987088; PID:9987105; PIDN:CAA60465.1  
C:Genetics:  
A:Gene: raph1  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

[illegible][illegible]

RESULT 8  
B35401  
cytochrome P450 105B1 - Streptomyces griseolus  
M:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Streptomyces griseolus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 04-Mar-2000  
C:Accession: B35401  
R:Omer, C.A.; Lentra, R.; Little, P.J.; Dean, C.; Tepperman, J.M.; Leto, K.J.; Romess  
J. Bacteriol. 172, 3335-3345, 1990  
A:Title: Genes for two herbicide-inducible cytochromes P-450 from Streptomyces griseo  
A:Reference number: A35401; MUID:90264332  
A:Accession: B35401  
A:Molecule type: DNA  
A:Residues: 1-403 <OME>  
A:Cross-references: GB:M32239; NID:g153480; PIDN:AAA26825.1; PID:g153481; GB:M36481  
C:GeneInfo:  
A:Gene: CYP105B1  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: Chromoprotein; electron transfer; heme; Iron; metalloprotein; monooxygenase  
E:238-374/Domain: cytochrome P450 homology <P45>  
F:352/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	15.6%	Score 340	DB 2	Length 403
Best Local Similarity	28.9%	Pred. No. 1.5e+18		
Matches 105	Conservative 63	Mismatches 165	Indels 30	Gaps 13

QY 63 MIATGGQILREAYEDYERHFSSSEC-----PTIPRENGEAYDIFPT--SMDPEQORFAL 115

Db 52 WLVTRHQVRAVLGDPR-FSADAHRTGFPFLTAGREIIGTNPFLRMDDPEHARLRML 110

QY 116 NOVGMAPVVDKLENIQGLACSLIESLRP-QGOCNFTEDYAEPPRIFFMLLAGPEEDI 174

Db 111 TADFTVKKVEAMRPVQGLADDLVDRMTTGRSADLVTEFALPLPSLYICLLIGVPIYDH 170

QY 175 PHL----KYLTDMTRPDGSMTPAEAKALYDYLPIIEORRQKPTDAIS-IYANGOVN 229

Db 171 AFQERSEVLLTLTRSTPE---EVRAQDELLEYLARLARTRKRRERDDAISRLVARGELD 227

QY 230 GRPITSDAKRMKCGLLLVGDLTYVNFSPSMELAKSPENRQELIEPEKIPAAECGL 289

Db 228 DTQIAT-----MGRLLVAGHEHTANMTALSTVLYLRNPDLARLARAPALVKGAVEELL 282

QY 290 RRFSLVADG--RIILSDYEFHGVOLKKGQJILLPOMLSGLDERKNACP--RHWDFSRQV 345

Db 283 RYLITVNHGVRINATEDVLIGRTIAGEGVLC--MISSARDAEVFPFGGDDLVARAR 340

QY 346 SHTTFGHSHCLCGOHLARREIIVTLKEMLTRIDPFSIA-PGAOIQHKSGI-VSGVALP 403

Db 341 RHVAFGFGVHOCGLQPLARVELQIAIEFLRLRLDLRLAVPHEEIPFGDMAIYGVHSLP 400

QY 404 LVM 406

Db 401 IAW 403

RESULT 9

F70729

Cytochrome P450 RV2266 - Mycobacterium tuberculosis (strain H37Rv)

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Mycobacterium tuberculosis

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: F70729

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

R:Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,

R:Jandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.



```

Db      54 QALLD-----PRIKDEFRTQOIIIEKRLADAERRPGSPDLGPIMLNTDPDPHTRLR 105
      113 ALANOVGMVYDKLENRIOLACSLIESLRPOGOCNFTEDYAEPPIRIF 172
      106 KLVKATARARVEELRRRIEDITDDLRLAGKSEVDIDFAFPLPTTYSLEMGVEDS 165
      173 DIPHLKYLTDQMTPRDPSMTFAEAK--ALYDYLPIIEORRQPGTDAISIVANGVNG 230
      166 RRDFRSMTNVLV--DGSOPBAQAASVAMVEYLTELAKRTPEPGDILLLEAVEEDG 223
      231 RPTISDAKRRKCGLLVGGDVTYVNFISFSEFLAKSPENHOELIERPERIPACEELR 290
      224 DRLSEGLIMAVFLLVAGHETVNLGNCLSLGNPDQALARNPDSLPGAIETELR 283
      291 RFLSVAGC--RLITSDFEFHGVOLKGDQILLPQMLSGLDERKNACPMHVDSPKOVSH 348
      284 YESVYANGTRHRTAEAVRFGDVVPEBELVWVAGLAANRDEGEREDPRDITRETTG 343
      349 TFGHSHLCLGOHLARREIIVTLKEMLTRIPDFSIA--PGAQIOHKSQIVSGVALPL 404
      344 AFGCHIFCVCAALRLAEQAIVAGRLLERPDLRMASPDLLRMRFVLMNGLEKLV 401

```

## RESULT 12

E69611

cytochrome P450 CYP4 - Bacillus subtilis

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Bacillus subtilis

C&gt;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 18-Feb-2000

C/Accession: E69611; T44774

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Berset

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd

A.; Ehlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gallier

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koelter, P.; Konungstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,

A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron

T.; Wilters, P.; Wipat, A.; Yamano, H.; Yamane, K.; Yasunoh, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033

A:Accession: E69611

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-410 &lt;KUN&gt;

A/Cross-references: GB:299117; GB:AL009126; NID:g2634966; PIDN:CAB14615.1; PID:el183903;

A:Experimental source: strain 168

R:Belitsky, B.R.; Gustafson, M.C.U.; Sonenshein, A.L.; von Wachenfeldt, C.

J. Bacteriol. 179, 5448-5457, 1997

A:Title: An lrp-like gene of Bacillus subtilis involved in branched-chain amino acid tra

A:Reference number: Z22837

A:Accession: T44774

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-410 &lt;BEL&gt;

A/Cross-references: EMBL:Y11043; PIDN:CAA71937.1

A:Experimental source: strain 1A1

C:Genetics: cypA

C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C:Keywords: oxidoreductase

F:245-381/Domain: cytochrome P450 homology &lt;CYP&gt;

```

QY      104 DPEOROFALANOVGMVYDKLENRIOLACSLIESLRPOGOCNFTEDYAEPPIRIF 163
      100 DPEPHNLRLTYLQCAFHNHMLIOLEDKIQHIALDSLKQVKNFMNLYVDYAFPLTIVI 159
      164 MLGAPDEEDIPHLKYLTDMQTRPDGSMTPAEAKL-----YDYLPIIEORQ 213
      160 SEMIGIPLEDRKFRVWSQA-----IDFSAPERLOENDHLLGEFEVELESIVRKKR 213
      214 KPGTDAISIVANGVNRPTISDEAKRMCGLLVGGDVTYVNFISFSEFLAKSPENHO 273
      214 EPAGDLISALIQASESTQSLTEELYSIMILLYAGHETVNLITNVTALMCHDQLEK 273
      274 LIERPERIPACEELLRFSLV--ADGRILTSDFEFHGVOLKGDQILLPQMLSGLDERK 331
      274 LRQOPDLMNSALIEALRPHSPVELTTRMTAEPTILHGOELKRKDVIIISLASANREKI 333
      332 NACPMHVDSPKOVSHHTFGHSHLCLGOHLARREIIVTLKEMLTRIPDFSIAPGAQIOH 391
      334 FPNADIDIERKNRHHAFGHNHFCIGAQDLRLAKIAISTLLRCPNIQL-KGEKKOM 392
      392 K--SGIVSGVALPL 404
      393 KWKGNFLMALIEELPI 408

```

## RESULT 13

A35401

cytochrome P450 105A1 - Streptomyces griseolus

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Streptomyces griseolus

C&gt;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 04-Mar-2000

R:Omer, C.A.; Lenstra, R.; Little, P.J.; Dean, C.; Tepperman, J.M.; Leto, K.J.; Romess

J. Bacteriol. 172, 3335-3345, 1990

A:Title: Genes for two hemicidic-inducible cytochromes P-450 from Streptomyces griseo

A:Reference number: A35401; MUID:90264332

A:Molecule type: DNA

A:Residues: 1-406 &lt;OME&gt;

A/Cross-references: GB:M32238; NID:g153477; PIDN:AAA26823.1; PID:g153478; GB:M36480

C:Genetics: cyp105A1

A:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygena

F:241-377/Domain: cytochrome P450 homology &lt;P45&gt;

F:355/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 15.1%; Score 330; DB 2; Length 406;

Best Local Similarity 26.3%; Pred. No. 8.9e-18;

Matches 104; Conservative 74; Mismatches 193; Indels 24; Gaps 11;

```

QY      31 PSNLISAGVQ--EAMAVIQESNVDPDLWTRCNGH-WIATGQLIREAYEDYR----- 79
      17 PSNNSCPYQPDGAQDLRDPGLRHRTLYDVGQAVVWTHNEAARKLLGDRPLSSNRTD 76
      80 HFSESECFIP--REAGAYVFIPTSMQPREQORFALANOVGMVYDKLENRIOLACSL 137
      77 NFPAVTSRFEAVRESPAF---IGLDPREHGRTRRTTISEFTVKRRKGMRPEVEVNG 132
      138 LIESLRPOG--QCNFTEDYAEPPIRIFMPLLAGPEEDIPHLKYLTDMQTRPDGSMTPAE 196
      133 FLDEMLAAGPTADLVDSGFALPVSMTYCRLLGYPADHFEFQDASKRVYSTQAQSLTA 192
      197 KEALYDYLPIIEORRQPGTDAI-SIVANGVNRPTISDEAKRMCGLLVGGDVTYVNF 255
      193 RNDLAGYLDGLITOFTEPEAGLVGALVADQLANG--IDREELISTPMILLINGHETTA 251
      256 FLFSMEFLAKSPENHOELIERPERIPACEELLRSL--VADGRILTSDFEFHGVOLK 313
      252 MTSLSVITLLDHPQYAAALRADRSLVPGAVEELLRYLAADIAGRGVATDIEVGHILR 311

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 13:04:51 ; Search time 27.02 Seconds

(without alignments)  
474.815 Million cell updates/sec

Title: US-09-246-451-11

Perfect score: 2180

Sequence: 1 TTETIQSNANLAPLPPHVE.....IVSGVALPLVMDPATTKAV 414

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2176	99.8	414	1	CPXA_PSEPU
2	591.5	27.1	414	1	Y4VG_RHISN
3	377	17.3	393	1	YH85_MYCTU
4	372	17.1	410	1	CPXM_BACME
5	366	16.8	396	1	YJIB_BACSU
6	361.5	16.6	405	1	CPXM_BACSU
7	360	16.5	398	1	Y218_MYCTU
8	340	15.6	402	1	CPXE_STRGO
9	338.5	15.5	428	1	YM66_MYCTU
10	335.5	15.4	410	1	CPXI_BACME
11	333.5	15.3	405	1	CPXK_SACER
12	333	15.3	410	1	CPXI_BACSU
13	330	15.1	405	1	CPXE_STRGO
14	319.5	14.7	405	1	YC56_MYCTU
15	318	14.6	428	1	CPXL_PSEPU
16	316.5	14.5	395	1	BIOL_BACSU
17	316.5	14.5	399	1	FAST_RHOFA
18	316	14.5	412	1	CPXH_STRGR
19	304	13.9	438	1	Y180_MYCTU
20	303	13.9	408	1	NOR_FUSOX
21	296.5	13.6	408	1	NOR2_CYLTO
22	295	13.5	381	1	CPXG_STRSO
23	291.5	13.4	433	1	Y245_MYCTU
24	288.5	13.2	403	1	NORI_CYLTO
25	288.5	13.2	404	1	CPXJ_SACER
26	277.5	12.7	397	1	CPXO_SACER
27	265	12.2	400	1	CPXP_RHISN
28	265	12.2	414	1	P778_MYCTU
29	253	11.6	401	1	CPXP_BRAJA
30	236	10.8	402	1	Y766_MYCTU
31	234	10.7	396	1	YM76_MYCTU
32	225	10.3	436	1	THCB_RHOER
33	220.5	10.1	400	1	YV21_MYCTU

34	217	10.0	422	1	CPXC_AGR6	P2466 agrobacteri
35	215	9.9	429	1	CPXR_BRAJA	O59204 bradyrhizob
36	206.5	9.5	489	1	YM68_MYCTU	O59572 mycobacteri
37	195	8.9	405	1	CYPX_BACSU	P34926 bacillus su
38	189	8.7	467	1	CPXR_RHISN	P55543 rhizobium s
39	177	8.1	368	1	CPXM_SULSO	O35080 sulfolobus
40	162	7.4	1048	1	CPXB_BACME	P14779 bacillus me
41	156.5	7.2	503	1	CP30_SHEEP	O29456 ovis aries
42	154.5	7.1	531	1	CP27_HUMAN	O02318 h cytochrom
43	151	6.9	407	1	CPXD_AGR6	P24667 agrobacteri
44	151	6.9	492	1	CPS3_PIG	O02390 sus scrofa
45	145.5	6.7	489	1	CPC3_RABIT	P00182 oryctolagus

## ALIGNMENTS

RESULT	1	ALIGNMENTS
CPXA_PSEPU	STANDARD;	PRT; 414 AA.
ID	CPXA_PSEPU	
AC	P00183:	
DT	21-JUL-1986 (Rel. 01, Created)	
DT	13-AUG-1987 (Rel. 05, Last sequence update)	
DT	15-FEB-2000 (Rel. 39, Last annotation update)	
DE	CYTCHROME P450-CAM (EC 1.14.15.1) (CAMPHOR 5-MONOOXYGENASE).	
GN	CAMC OR CYP101.	
OS	Pseudomonas putida.	
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;	
OC	Pseudomonas.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE: 8611751.	
RA	Unger B.P., Gunsalus I.C., Sligar S.G.;	
RT	"Nucleotide sequence of the Pseudomonas putida cytochrome P-450cam	
RT	gene and its expression in Escherichia coli.";	
RL	J. Biol. Chem. 261:1158-1163(1986).	
RN	[2]	
RP	SEQUENCE OF 385-414 FROM N.A.	
RP	SPRAIN-ATCC 17453;	
RC	MEDLINE: 90130389.	
RX	MEDLINE: 90130389.	
RA	Koga H., Yamaguchi E., Matsunaga K., Aramaki H., Horiuchi T.;	
RT	"Cloning and nucleotide sequences of NADH-putidaredoxin reductase	
RT	P-450cam and putidaredoxin gene (cam) involved in cytochrome	
RT	P-450cam hydroxylase of Pseudomonas putida.";	
RL	J. Biochem. 106:831-836(1989).	
RN	[3]	
RP	SEQUENCE.	
RX	MEDLINE: 83030788.	
RA	Hanlu M., Arnes L.G., Yasunobu K.T., Shastri B.A., Gunsalus I.C.;	
RT	"Amino acid sequence of the Pseudomonas putida cytochrome P-450. II.	
RT	Cytogen bromide peptides, acid cleavage peptides, and the complete	
RT	sequence.";	
RL	J. Biol. Chem. 257:12664-12671(1982).	
RN	[4]	
RP	X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).	
RX	MEDLINE: 86143817.	
RA	Finzel B.C., Weber P.C., Hardman K.D., Salameh F.R.;	
RT	"Structure of ferricytochrome c' from Rhodospirillum rubrum at	
RT	1.67-A resolution.";	
RL	J. Mol. Biol. 186:627-643(1985).	
RN	[5]	
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).	
RX	MEDLINE: 86059514.	
RA	Poulos T.U., Finzel B.C., Gunsalus I.C., Wagner G.C., Kraut J.;	
RT	"The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450.";	
RL	J. Biol. Chem. 260:16122-16130(1985).	
RN	[6]	
RP	X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).	
RX	MEDLINE: 98019009.	
RA	Schlichting I., Jung C., Schulze H.;	
RT	"Crystal structure of cytochrome P-450cam complexed with the (1S)-	
RT	camphor enantiomer.";	
RL	FEBS Lett. 415:253-257(1997).	

```

RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE; 98313255.
RA Vlakovic M., Silgar S.G., Li H., Poulos T.L.;
RT "Understanding the role of the essential Asp251 in cytochrome p450cam
   using site-directed mutagenesis, crystallography, and kinetic solvent
   isotope effect.";
RL Biochemistry 37:9211-9219(1998).
RN [8]
RP STRUCTURE BY NMR.
RX MEDLINE; 97459726.
RA Mourou C., Bondon A., Stimmoneaux G., Jung C.;
RT "H-NMR study of diamagnetic cytochrome p450cam: assignment of heme
   resonances and substrate dependence of one cysteinate beta proton.";
RL FEBS Lett. 414:203-208(1997).
CC -1- FUNCTION: INVOLVED IN A CAMPHOR OXIDATION SYSTEM.
CC -1- CATALYTIC ACTIVITY: CAMPHOR + PUTIDAREDOXIN + O(2) = 5-EXO-
CC HYDROXYCAMPHOR + OXIDIZED PUTIDAREDOXIN + H(2O).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; M12546; AAA25760.1; -.
DR EMBL; D00528; BAA00412.1; -.
PIR; A25660; OAPSCP.
DR PDB; 2CPR; 1S-APR-91.
DR PDB; 3CPP; 1S-APR-91.
DR PDB; 4CPP; 1S-JUL-91.
DR PDB; 5CPP; 1S-JUL-91.
DR PDB; 6CPP; 1S-JUL-91.
DR PDB; 7CPP; 1S-JUL-91.
DR PDB; 8CPP; 1S-JUL-91.
DR PDB; 1CP4; 1S-JAN-93.
DR PDB; 2CP4; 1S-JAN-93.
DR PDB; 3CP4; 1S-JAN-93.
DR PDB; 4CP4; 1S-JAN-93.
DR PDB; 5CP4; 1S-SEP-98.
DR PDB; 6CP4; 1S-SEP-98.
DR PDB; 1NOO; 08-MAR-96.
DR PDB; 1PHA; 31-OCT-93.
DR PDB; 1PHB; 31-OCT-93.
DR PDB; 1PHC; 31-OCT-93.
DR PDB; 1PHD; 31-OCT-93.
DR PDB; 1PHE; 31-OCT-93.
DR PDB; 1PHF; 31-OCT-93.
DR PDB; 1PHG; 31-OCT-93.
DR PDB; 1AKD; 19-NOV-97.
DR PFAM; PF00067; P450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
   3D-structure.
FT INT_MET 0 0
FT BINDING 357 357 HEME.
FT CONFLICT 55 56 MISSING (IN REF. 3).
FT CONFLICT 276 276 E -> Q (IN REF. 3).
FT CONFLICT 361 361 H -> S (IN REF. 3).
FT CONFLICT 407 407 D -> N (IN REF. 3).
FT TURN 16 17
FT HELIX 20 22
FT STRAND 23 23
FT TURN 28 29
FT TURN 32 33
FT HELIX 34 36
FT HELIX 38 46
FT TURN 48 49
FT STRAND 53 56
FT HELIX 58 60

```

FT	STRAND	62	65
FT	HELIX	68	76
FT	TURN	78	80
FT	STRAND	81	82
FT	TURN	83	84
FT	HELIX	90	95
FT	TURN	99	102
FT	TURN	105	107
FT	HELIX	108	119
FT	HELIX	121	145
FT	TURN	146	146
FT	STRAND	147	149
FT	HELIX	150	153
FT	TURN	154	156
FT	HELIX	157	167
FT	TURN	168	168
FT	HELIX	171	185
FT	HELIX	193	213
FT	HELIX	219	224
FT	TURN	225	225
FT	STRAND	227	228
FT	TURN	229	230
FT	STRAND	231	232
FT	HELIX	235	250
FT	TURN	251	252
FT	HELIX	253	266
FT	HELIX	268	276
FT	HELIX	278	291
FT	TURN	292	292
FT	STRAND	295	295
FT	STRAND	297	301
FT	STRAND	305	307
FT	TURN	308	309
FT	STRAND	310	312
FT	TURN	314	315
FT	STRAND	317	320
FT	HELIX	322	324
FT	TURN	325	327
FT	TURN	329	331
FT	TURN	335	336
FT	TURN	340	341
FT	TURN	349	350
FT	HELIX	353	355
FT	TURN	358	359
FT	HELIX	360	377
FT	STRAND	382	383
FT	TURN	385	386
FT	STRAND	391	392
FT	STRAND	396	396
FT	STRAND	398	399
FT	STRAND	403	405
FT	HELIX	408	410
SO	SEQUENCE	414 AA;	46538 MW;
			ECAG10293A9D6207 CRC64;

Query Match	Similarity	99.8%	Score	2176:	DB 1:	Length	414:
Best Local	Similarity	99.8%	Pred. No.	9e-159:			
Matches	413:	Conservative	1:	Mismatches	0:	Indels	0:
QY	1	TTTGTIOSANANLALP	PHVPEHLV	PFDDMYNPSNL	SAGVOEAMAVLQES	NPDLVTRCNG	60
Db	1	TTTGTIOSANANLALP	PHVPEHLV	PFDDMYNPSNL	SAGVOEAMAVLQES	NPDLVTRCNG	60
QY	61	GMWITRQOLIREA	EDYRHFSECP	FTPREGGEAYD	FPTMDPPEOROFAL	NOVYG	120
Db	61	GMWITRQOLIREA	EDYRHFSECP	FTPREGGEAYD	FPTMDPPEOROFAL	NOVYG	120
QY	121	MPVYKLENRIOL	ELACSLTESLR	POGQCNFTEDVA	AEPPPIRIFML	LAGLPEEDIFHL	180
Db	121	MPVYKLENRIOL	ELACSLTESLR	POGQCNFTEDVA	AEPPPIRIFML	LAGLPEEDIFHL	180
QY	181	TDQDMRPDGSMT	FAEAKALYD	LPIITEOROKAGTA	ISIVANGQVNGR	ITISDEAKR	240

Db	181	TDQMRPDGSMTFPEAKREALDYLIPIEQROKRPQTDAISIVANGVGNRPITSDEAKR	240
Qy	241	MGGLLVGGLDVTYVNFISFSMEFLAKSPDRHÖLIERPERIIPAACELLRRFSIVADGRI	3000
Db	241	MGGLLVGGLDVTYVNFISFSMEFLAKSPDRHÖLIERPERIIPAACELLRRFSIVADGRI	3000
Qy	301	LTSYEFHGVLKKGGDQIILLPQMLSGDERKKNCPMIVNDESRÖKVSHITTTGHGSHLCLGQ	3600
Db	301	LTSYEFHGVLKKGGDQIILLPQMLSGDERKKNCPMIVNDESRÖKVSHITTTGHGSHLCLGQ	3600
Qy	361	HLAREEIVTLKEWLTRIPFESIAPGAQIOHKSGIYGVQALPLVMPDATTTKAV	414
Db	361	HLAREEIVTLKEWLTRIPFESIAPGAQIOHKSGIYGVQALPLVMPDATTTKAV	414

RESULT	2			
YAVG_RHISN	YAVG_RHISN	STANDARD:	PRT:	414 AA.
AC	053215;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	PROBABLE CYTOCHROME P450 YAVG (EC 1.14.14.-).			
GN	YAVG.			
OS	Rhizobium sp. (strain NGR234).			
OG	Plasmid sym pNGR234a.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;			
OC	Rhizobiaceae; Rhizobium.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 97305956.			
RA	Freiberg C.A., Fellay R., Balroch A., Broughton W.J., Rosenthal A.,			
RA	Perret X.;			
RT	"Molecular basis of symbiosis between Rhizobium and legumes.";			
RL	Nature 387:394-401(1997).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 96389014.			
RA	Freiberg C., Perret X., Broughton W.J., Rosenthal A.;			
RT	"Sequencing the 500-kb GC-rich symbiotic replicon of Rhizobium sp.			
RT	NGR234 using dye terminators and a thermostable 'sequenase': a			
RT	beginning.";			
RL	Genome Res. 6:590-600(1996).			
CC	-I- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE			
CC	MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED			
CC	COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.			
CC	-I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/isb-sib.ch">http://www.isb-sib.ch/announce/isb-sib.ch</a> ).			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
EMBL	Z68203; CA92422.1; -			
DR	EMBL; AE000101; AAB91895.1; -			
DR	HSSP; P00183; 3CP4.			
DR	PFAM; PF00067; P450; 2.			
DR	PROSITE; PS00086; CYTOCHROME_P450; FALSE_NEG.			
KM	Hypothetical protein: Oxidoreductase; Monooxygenase;			
KM	Electron transport; Heme; Plasmid.			
FT	BINDING 364 364 HEME (BY SIMILARITY)			
FT	SEQUENCE 414 AA; 45810 MW; 888598BE99315BB84 CXC64;			

Query Match	27.1%	Score 591.5	DB.1	Length 414
Best Local Similarity	34.1%	Prod.No. 5.9e-38		
Matches 142: Conservative	65	Mismatches 171	Indels 39	Gaps 9
13 PLPRVPRHLYEDPMYPSNLSA---GVQEMANVLQESNNPDVW-----TRCNGHWIA 65				

[illegible]

RESULT	3
YH85_MYCTU	
ID	YH85_MYCTU
AC	053936
DT	15-FEB-2000 (Rel. 39, Created)
DT	15-FEB-2000 (Rel. 39, Last sequence update)
DT	15-FEB-2000 (Rel. 39, Last annotation update)
DE	PUPATIVE CYTOCHROME P450 RV1785C (EC 1.14.-.-).
GN	RV1785C OR MTW049.07C.
OS	Mycobacterium tuberculosis
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-H37RV;
RX	MEDLINE: 98295987.
RA	Cole S.T., Brosch R., Parkhill J., Garner T., Churcher C., Harris D.,
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA	Bacckock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Dajugs K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Hornsby T., Javels K., Krogsh A., McLean J., Moule S., Murphy L.,
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA	Taylor K., Whitehead S., Barrall B.G.;
RT	"Deciphering the biology of Mycobacterium tuberculosis from the
RL	complete genome sequence."
RL	Nature 393:537-544(1998).
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL: AL022021; CAAL1707.1; -
DR	TUBERCULIST: RV1785C; -
DR	PFAM: PF00067; P450; 2.
DR	PROSITE: PS00086; CYTOCHROME_P450; 1.
FT	Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
FT	BINDING 342 342 HEME (BY SIMILARITY).
FT	SEQUENCE 393 AA; 43541 MW; 8BCDDID508471FEAB8 CXC64;

DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
 FT BINDING 342 HEME (BY SIMILARITY).  
 SO SEQUENCE 393 AA; 43541 MW; 8BCDDID508477FAB8 CnC64;

Query Match	17.3%	Score 377;	DB 1;	Length 393;
Local Similarity	26.7%;	Pred. No. 1.3e-21;		
Matches 110;	Conservative 70;	Mismatches 183;	Indels 20;	Gaps 10;
QY	31	PSNLASGVQAEAMAVLQESNVDPDVTWTCNGGHMIAITRGQLIREAVEDYKHFSSECPFIIPR	90	
Db	23	PMAYDGVG--GKTLTRDAG--PVVE--MNGWYLL--TRRDYLAALRNPKVFSSKALQP-	74	
QY	91	EAGEYVDITPIPTSMDEPQORFALANQVGMVYDKLENRIODELACSLTESLRPGQCNE	150	
Db	75	-PGNDLPVPLAFDPEPETHRYRRIIQPYFSSPALSKALPSLRHRTVAMIDALNAGECEA	133	
QY	151	TEDVPEPPRIREFMLLAGIPREDITHLNYLDQM----TRPGSMTFPAEAKALDYLIIP	206	
Db	134	MADLANLFFPQLEVLVYGLPREDRRLIGWKAYVATMSDRPHTADVAARELLEYLTA	193	
QY	207	ITEORRQKPGTDAISIVANGQVNGRPITSDEAKRMCGLLVGLDTPVNFLESFSEMTAK	266	
Db	194	MVAEHRNRNPGPVLQV---QIGEDPLSEIVLGLSHLLIAGDQTVAAVAFSILELAR	250	
QY	267	SPEHROELIERPERIPAAECLELR-RFSLVADGRILTSDFEFGVQLKGGDILLPQMLS	325	
Db	251	RPQLRAMLRDNPQKQIRVFEIEIVRLPESPAPVAPRVTEPVTGVMTLPGSVRLCMAAV	310	
QY	326	GLDEKKNACPMHVDSSRQKVSHTTGHSHSLCGLHARREITVILKELRIIPDFSIAP	385	
Db	311	NRDGSAMDSTDELVDMDGVYHRHMGGGGPHRCGLSHARLETLLVGEMLIDPDEIAR	370	
QY	386	--GAQIOHKSQIVSGVQALPLWV	406	
Db	371	DYAPEIRPSSKFA--LKNLPLRW	392	
RESULT	4			
CPXM_BACME				
ID	CPXM_BACME	STANDARD:	PRT:	410 AA.
AC	006069;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	CYCLOCHROME P450(MEG) (EC 1.14.99.-) (STEROID 15-BETA-HYDROXYLASE)			
DE	(STEROID 15-BETA-MONOOXYGENASE).			
GN	CYP106A2.			
OC	Bacillus megaterium.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN-ATCC 13368;			
RA	MEDLINE: 94049677.			
RA	Rauschenbach R., Isernhagen M., Noeske-Jungblut C., Boldol W.,			
RA	Siewert G.;			
RT	"Cloning, sequencing and expression of the gene for cytochrome			
RT	P450meg, the steroid-15 beta-monooxygenase from Bacillus megaterium			
RT	ATCC 13368."			
RL	Mol. Gen. Genet. 241:170-176(1993).			
RN	[2]			
RP	CHARACTERIZATION.			
RC	STRAIN-ATCC 13368;			
RC	MEDLINE: 79194051.			
RA	Berg A., Ingelman-Sundberg M., Gustafsson M.;			
RT	"Purification and characterization of cytochrome P-450meg."			
RT	J. Biol. Chem. 254:5264-5271(1979).			
RN	[3]			
RP	CHARACTERIZATION.			
RC	STRAIN-ATCC 13368;			
RC	MEDLINE: 82091079.			
RA	Berg A., Rafter J.J.;			
RT	"Studies on the substrate specificity and inducibility of cytochrome			
RT	P-450meg."			
RL	Biochem. J. 196:781-786(1981).			
CC	-1- FUNCTION: HAS THE CAPACITY TO HYDROXYLATE CERTAIN STEROIDS IN THE			
CC	15-BETA POSITION.			

[illegible]

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF015825; AAC46317.1; -.  
DR EMBL: Z99110; CAB13078.1; -.  
DR SUBMITTER: BG13195; YJIB.  
DR PEAM: PF00067; P450.1.  
DR PRINTS: PR00359; BP450.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
FT BINDING 349 349 HEME (BY SIMILARITY).  
SQ SEQUENCE 396 AA; 44990 MW; 9A89CF12613DBCFB CRC64;  
  
Query Match 16.8%; Score 366; DB 1; Length 396;  
Best Local Similarity 30.4%; Pred. No. 8.8e-21;  
Matches 100; Conservative 54; Mismatches 157; Indels 18; Gaps 7;  
  
OY 71 IREAYEDYRHHSSCPFIPREAGEAVDFIPMSDPPREORORFALANOVGVVYDKLENR 130  
Db 56 VKKVVGKDELFSS-C--MPQOTSSIGNST-IMDPKHTKIRSYNKKFTPRVMKQWEP 111  
OY 131 IOELACSLIESLRPOGOCNFTEDYAEPPIRIFMLAGLPREDIPHLKYLTDOMT----- 185  
Db 112 IOETDELIOKFGQRSEEDLVHDSYPLVIVISELIGVPSAHMEOFKAMBDLVSTPKD 171  
OY 186 -RPGSMFAE-----AKALYDYLPIITEQRKRGTDALSTIVANGVNGPRTSDEAKR 240  
Db 172 KSEAEKNAFLERDKCEBELAFAFAGIIEEKRNRKPEODIISLVEAETGKLSGEELIP 231  
OY 241 MCGLLVAGLTVVNFPSMEPLAKSPENHROLIERPERIPACEELLR-RFSIVACGR 299  
Db 232 FCTLLVAGNFTTNLSNANWSTLETPGYVELSHPELPOAVEELRAPAPVLR 291  
OY 300 ILTSDYEHGVOLKKGDOILLPOMLSGLDERKNACPMHVDSEKQVSHTTGGSHLCLG 359  
Db 292 IAKDTEIGHLIKEGDVLAFVANSANDEAKPRPHMFDLRHNPRIIAGHGHCFLG 351  
OY 360 QHLAREIIVLKEMLTRIPD-----STAP 385  
Db 352 APLARLEANIATLSLISAFPHMECVSTIP 380  
  
RESULT 6  
CPXM\_BACSU STANDARD; PRT; 405 AA.  
AC P27632;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE CYTOCHROME P450 109 (EC 1.14.-.-) (ORF405).  
GN CYP109.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
CC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-W23;  
RX MEDLINE; 91192601.  
RA Ahn K.S., Wake R.G.;  
RT "Variations and coding features of the sequence spanning the  
RT replication terminus of Bacillus subtilis 168 and W23 chromosomes.";  
RL Gene 98:107-112(1991).  
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLEATE  
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED

CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M24523; AAA22720.1; -.  
DR HSSP: Q00441; IOXA.  
DR PEAM: PF00067; P450.1.  
DR PRINTS: PR00359; BP450.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Heme.  
FT BINDING 351 351 HEME (BY SIMILARITY).  
SQ SEQUENCE 405 AA; 45845 MW; 1331D5BEA7E3C04 CRC64;  
  
Query Match 16.6%; Score 361.5; DB 1; Length 405;  
Best Local Similarity 25.2%; Pred. No. 2e-20;  
Matches 113; Conservative 72; Mismatches 169; Indels 95; Gaps 10;  
  
OY 1 TTETIOSN-----ANLALPP-HYPEHLVFPEDMYNPSNLASGV-----OEAMAVLOESN 49  
Db 2 TNOTARSSKKRRYANLIPMEELHSEKRLFPPIYDKLRSPYRHYDPLRCMDVFK--- 58  
OY 50 VPDLVWTRCNGHMIATRGOLIREAYEDYRHHSSCPFIPREAGEAVDFIPTSDPPEOR 109  
Db 59 -----YDDQVFVKNPKLFSSKRGIQTESILT-MDPKHT 92  
OY 110 QPRALANOVGVVYDKLENRIQELACSLIESLRPOGOCNFTEDYAEPPIRIFMLAGL 169  
Db 93 KLRLVSAFPTPKAVKQLETRIKIDVTAFILOEAROKSTIDIEDFAGPLPVIIIAEMIGA 152  
OY 170 PEEDIPHLKYLTDOMTRDGSMTFAEKA-----LYYLLPIIEOR 211  
Db 153 PIEDRHILIKITYSDLV-----AGAKDSSDKAVADVWNRROCHAFSLFRDITLKR 204  
OY 212 RQKGTDAISIVANGVNGRPITSDEAKRMGLLVGLDITVNFPSMEPLAKSPENR 271  
Db 205 RAEPKEDMTMLQAEIDEGYLTQEQLGFCILLVACNETTNNLAAVAVYLTEDSVYQ 264  
OY 272 QELIERPERIPACEELLRESLV-ADGRILITSDYEHGVOLKKGDOILLPOMLSGLDER 330  
Db 265 QOVRONTDNVANVIEETLRYSPOVAIGRVATEDEDELGVEFIKGSYSVIASANRDED 324  
OY 331 KNACPMHVDSEKQVSHTTGGSHLCLGQHLAREIIVTLK-----EML 375  
Db 325 KFKRPDCKRTIPSPYSPHLSFGFGIHFCLGAPLARLEANIATLSLISMSACIEKAHDEKL 384  
OY 376 TRIPDFSIAPGAIOHKSGIVGVOALPL 404  
Db 385 EALP-----SPFVGVKRLPV 400  
  
RESULT 7  
Y218\_MYCTU STANDARD; PRT; 398 AA.  
AC O53563;  
DT 15-FEB-2000 (Rel. 39, Created)  
DT 15-FEB-2000 (Rel. 39, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PUTATIVE CYTOCHROME P450 RV3518C (EC 1.14.-.-).  
GN Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;

RX MEDLINE: 98295987.  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter K., Seeger K., Skelton S., Squares S., Rogers J.,  
 RA Taylor K., Whitehead S., Barrall B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AL022022; CA17755.1; -.  
 DR TUBERCULIST; RV3518C; -.  
 DR PRAM; PF00067; P450; 1.  
 DR PRINTS; PR00359; BP450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KM Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
 FT BINDING 340  
 FT SEQUENCE 398 AA; 44398 MW; BCF3C23ECB5767F CRC64;

Query Match 16.5%; Score 360; DB 1; Length 398;  
 Best Local Similarity 28.2%; Pred. No. 2,5e-20;  
 Matches 107; Conservative 69; Mismatches 169; Indels 34; Gaps 8;

OY 55 WTRCN-----GGHWIATRGQIREAYEDYRHFS-----ECPIPREAGEAYDPIPT 101  
 DB 26 WMRANQVFRNRNGLAASTYQAVIDAERQPELFSNAGIRPDQPALP-----MMI 76  
 OY 102 SMDPEQRORPALANOVGVNVDKLENRTQELACSLIESRPGCCNFTEDYAEPPPIR 161  
 DB 77 DMDPAHLRLKRLVNAAGFTTRKRVNDRKASIALCDTLIDACGECEDFVNDLAAPRMA 136  
 OY 162 IFMLAGL-PEEDIPLHK-----YLTDMTRPDGSMTEFAEAKALYDYLPIPIEQRQ 213  
 DB 137 VIGMGLVREQRDMPFLRMWDLDVTLFSHVSOEDPOITM-DATAAINDPTRAIARRA 195  
 OY 214 KPGTDAISIVANGOVNCRPITSDEAKRMGILLVGLDVTYVNFSLFSMEPLAKSPEHROE 273  
 DB 196 DPTDVLVSIVSSVEVDEERLSDELVMETLLILIGDETTTRHFLSGTQDLRRNDQDL 255  
 OY 274 LIERPERIPACEELRLRFSIVAD-GRILTSDFEHGVOLKKGQIILLPQMLSLDERKN 332  
 DB 256 LQDRPSLLPGRIEMLMTWPAVKMKCVLTRADEFGHTALCAGKMKMLESANFEDAVP 315  
 OY 333 ACPMHVDFSRQKSHTTFGHSHLCLGOHLARREIIVTLKEMLRIPDPSIAPGAQI--Q 390  
 DB 316 CEPEKPEFVQGNPNSHLAFGFGTHFCGLGNQLARLELSMTERVRLRLPDLRLVADSVLPL 375  
 OY 391 HKSGIVSGVQALPLWMDPA 409  
 DB 376 RPAFNVSGLESMVVPFPPS 394

RESULT 8  
 CPXF\_STRGO STANDARD; PRT; 402 AA.  
 ID CPXF\_STRGO STANDARD; PRT; 402 AA.  
 AC P18327;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE CYTOCHROME P450-SU2 (EC 1.14.-.-) (P450-CVB1) (CYP105B1).

GN CYP105B1 OR SUBC.  
 OS Streptomyces griseolus.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomyicinae; Streptomycetaceae; Streptomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-32.  
 RC STRAIN-ATCC 11796;  
 RX MEDLINE: 90264332.  
 RA Omer C.A., Lemstra R., Little P.J., Dean C., Tepperman J.M.,  
 RA Lato K.J., Romesser J.A., O'Keefe D.P.;  
 RT "Genes for two herbicide-inducible cytochromes P-450 from  
 RT Streptomyces griseolus."  
 RL J. Bacteriol. 172:3335-3345(1990).  
 CC -1- FUNCTION: METABOLISM OF A NUMBER OF SULFONYLUREA HERBICIDES.  
 CC -1- INDUCTION: BY HERBICIDE.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M32239; AAA26825.1; -.  
 DR PIR; B35401; B35401.  
 DR HSSP; P23295; 2ROM.  
 DR PRAM; PF00067; P450; 1.  
 DR PRINTS; PR00359; BP450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KM Oxidoreductase; Monooxygenase; Electron transport; Heme.  
 FT INIT MET 0  
 FT BINDING 351 351 HEME (BY SIMILARITY).  
 FT SEQUENCE 402 AA; 44278 MW; E3B67F672C2609D CRC64;

Query Match 15.6%; Score 340; DB 1; Length 402;  
 Best Local Similarity 28.9%; Pred. No. 8,6e-19;  
 Matches 105; Conservative 63; Mismatches 165; Indels 30; Gaps 13;

OY 63 WINTRGQIREAYEDYRHFSSEC-----PIPREAGEAYDPIPT--SMDPEQRORPALA 115  
 DB 51 WLVTIRHODVRAVLGDPR-FSADAHRTGFPFLTAGGREIIGTNPFLRMDDPEHARLRML 109  
 OY 116 NOVGVNVDKLENRTQELACSLIESLR-QGOCNFTEDYAEPPPIRIFMLAGLPREDI 174  
 DB 110 TADFTYAKVEAMREYQRLADDVDRMTTGTSDVTEFALPLPSLVTLGLGVPYEDH 169  
 OY 175 PHL-----KYLTDMTRPDGSMTEFAEAKALYDYLPIPIEQRORPGTDAIS-IVANGOVN 229  
 DB 170 AFQGERSRVLLTLRSPE---EVRANODELLFYLARLARKRKRRPDALISRLVANGELD 226  
 OY 230 GRPITSDEAKRMGILLVGLDVTYVNFSLFSMEPLAKSPEHROELIERPRIPACEELL 289  
 DB 227 DTQIAT-----MGRLILVAAGHETTANMTALSTVLLRNPDQLRLRAEPALVAGAVEELL 281  
 OY 290 RRSIVADG--RLITSYERHGVOLKKGQIILLPQMLSGIDERKNCP--MHVDFSNQKV 345  
 DB 282 RYLLIVANGVRATATEVLLIGRTIAGEGVLC--MTSSANRDAEVPFGDDLDVADAR 339  
 OY 346 SHTTFGHSHLCLGOHLARREIIVTLKEMLRIPDPSI-PGAQIOHKSIGI-VSGVQALP 403  
 DB 340 RHVAFGCGVHQCIGQPLARVELQIATITLLRLPDLRLAVPHEIIPRGDMALTYGHSLP 399  
 OY 404 LVW 406  
 DB 400 IAW 402

RESULT 9  
 YM66\_MYCTU STANDARD; PRT; 428 AA.  
 ID YM66\_MYCTU STANDARD; PRT; 428 AA.

AC O50696;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PUTATIVE CYTOCHROME P450 RV2266 (EC 1.14.-.-).  
GN RV2266 OR MTCY339.44C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV.  
RX MEDLINE; 98295987.  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigelsmeier K., Gas S., Barry C.E. III, Tekaia F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Kitchin A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z717163; CAB00969.1; -.  
DR HSSP; P33006; 1CPT.  
DR TUBERCULIST; RV2266; -.  
DR PFAM; PF00067; P450.1.  
DR PRINTS; PR00359; BP450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; FALSE\_NEG.  
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
FT BINDING 379 379 HEME (BY SIMILARITY).  
SQ SEQUENCE 428 AA; 47824 MW; 76B1F3C5AE348591 CRC64;

Query Match 15.5%; Score 338.5; DB 1; Length 428;  
Best Local Similarity 26.2%; Pred. No. 1.2e-18;  
Matches 117; Conservative 72; Mismatches 188; Indels 69; Gaps 17;

QY 5 IQSANLAPLPHPVPHENLVFDFDMYNPSNLACVOEAMAVIQE-----SNVPLV 54  
DB 7 IATRYNGTP-PPEVP--IADIEL-----GSLDFWALDDVDGAFATLRREAPISF 54  
QY 55 WTR-----CNGGHMIATRCQLIREAEDY---RH---FSS-----ECPTIPREA 92  
DB 55 WPTIELPGFVANGNMALTK-----YDVFYASRHRPDISSYINNTITINQTPLEAYF 107  
QY 93 GEAYDFTISMDPEQROFRALANOVGMVYVDKLENRIQELACSLIESL---RPQGCN 149  
DB 108 GSM-----IVLDDPRHQRISIVSRAFTPKVARIETEAIVRDAHRLVSSMIANNPDRDAD 162  
QY 150 FTEDYAEFPPIRIFMLLAGLPREDIPHLAKYLTDMT---RPGDSMTFPAEAKALYD---Y 203  
DB 163 LVSELTAGLPQIICDDMGIRPADHQRIHMTNVLGFGDPLAVDFEFQVNSADIGAY 222  
QY 204 LPIITEORRORPGTDAISIVANGVNGRPITSDEAKRMCGLLVGLTVPVFLSFSMEF 263  
DB 223 ATALAEADRVRNHHDUTLSIVLEAEDGRSLSRLEIASFFILLVAGNETRTNATHGVLA 282  
QY 264 LAKSEHROELIERPPI-PAACEELLRRFS-LVADGRIILTSYDFYHGVLKKGQDILP 321  
DB 283 LSRYPEORDRWMSDFGLAPTAVEEIVWMAFPVYVVRRTLIQDILRGTAKMAAGKVSIM 342

QY 322 OMLGDERKNKAPMHVDFSRKOVSHSTTF-GHSGHLCGLHARREITVTLKEMLTRIPD 380  
DB 343 YCSANRDSKRADPTFTPLANPNPHLGFGGGAIFCGANLAREINVARDELAROMPD 402  
QY 361 FSIAPGAQIQHKSQIVSGVALPLVW 406  
DB 403 V-VAITEEPARLLSQFIHGIKTLPTVW 427

RESULT 10  
CPXL\_BACME  
ID CPXL\_BACME STANDARD; PRT; 410 AA.  
AC P14762;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME P450(BM-1) (EC 1.14.14.1).  
GN CYP106.  
OS Bacillus megaterium.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-25.  
RC STRAIN-ATCC 14581;  
RX MEDLINE; 90089408.  
RA He J.S., Ruettinger R.T., Liu H.-M., Fulco A.J.;  
RT "Molecular cloning, coding nucleotides and the deduced amino acid  
RT sequence of P-450BM-1 from Bacillus megaterium."  
RL Biochim. Biophys. Acta 1009:301-303(1989).  
RN [2]  
RP SEQUENCE OF 1-25 FROM N.A.  
RX MEDLINE; 95355495.  
RA He J.S., Liang Q., Fulco A.J.;  
RT "The molecular cloning and characterization of Bm1p1 and Bm1p2  
RT proteins, putative positive transcription factors involved in  
RT barbiturate-mediated induction of the genes encoding cytochrome  
RT P450BM-1 of Bacillus megaterium."  
RL J. Biol. Chem. 270:18615-18625(1995).  
CC -I- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED  
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.  
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X16610; CAA34612.1; -.  
DR EMBL; S79230; AAC60495.1; -.  
DR PIR; S07764; O4856M.  
DR HSSP; P33006; 1CPT.  
DR PFAM; PF00067; P450.1.  
DR PRINTS; PR00359; BP450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.  
FT BINDING 356 356 HEME.  
SQ SEQUENCE 410 AA; 47460 MW; C9AE293EF6745387 CRC64;

Query Match 15.4%; Score 335.5; DB 1; Length 410;  
Best Local Similarity 27.7%; Pred. No. 1.9e-18;  
Matches 96; Conservative 57; Mismatches 170; Indels 23; Gaps 7;

QY 63 WIATGQLIREYEDYRRFSSP---CPPIPRAGALYFIP---TSMDDPEQROFRALA 115  
DB 45 WNVFYEYHVKOVLNYSDFSSDGGRTTIFVGNSSKRKSTSTPTNLJTNDDPPHRRARSIL 104  
QY 116 NOVGMVYVDKLENRIQELACSLIESLRPQGCNTEDEYAEFPPIRIFMLLAGLPREDIP 175



Db	105	AAAFPRSLKMWEDRIKQIADLVLEAIQKNTINIVDDLSPPFSYIADLFGVPVADRY	164
Qy	176	HLKTYTDMQTRPDGSMTFAEAKE-----ALYDYLPIITIEQRQKPGTDAISVANGQV	228
Db	165	QEKRWVDILFPYPOERLEEIOEKORAGAEFYGLYPIVETEKRSNSDILISDLOAEV	224
Qy	229	NGRPITSDPEAKRMCGILLVGGDGVVNVLT-SFSMEPLAKSPENHQELIERPERIPACEE	287
Db	225	DGEFTTDEEYIHAATMLLGLAGAVETTSIIAMFYSFLYDOKSLYSELRRNDELAPKAVEE	284
Qy	288	LLR-RFSLVADGRILITSDDEFHGVQLAKKQDQILLPOMLSGLDERKKNACPMHVFDSR-QKV	345
Db	285	MLRYRFHISRDRIVYKQDNELLGVKLKKGDVYIAMSACNNDETFMENPESVDIHRPTNK	344
Qy	346	SHTTFGSHSLCTLQGLAHARRIYITLKEMLTR-----IPDFSIAP	385
Db	345	KHLTFGNGPHFLCGLAPLARLEMKITILEAFVFSHIEPFEDELELP	390

Db	54	QALLD-----PRINKDEFRTOOILEKRIADAERRPGFSPDLGPHMLTDPDHTLR	105
Qy	113	ALANQVGMVYDKLENIENIELASLIESLRPOGOCNTEDEVAEPPIRIFMLAGLPEE	172
Db	106	KLVKAFARVEGRRIEIQITDOLLRLAGREVDLIDEPAPPLIPIYISLMEGVDS	165
Qy	173	DIPHLKITTDMPDQSMFFAEKE--ALXDYILPIEORROKPGTDAISIVANGVNG	230
Db	166	RDDFERSTNVLV--DSOPEAOQAQAVAMVEYLTETIAKRREPGDOLLTALLEVEDG	223
Qy	231	RPIIIDEAKRMCGILLVGGDITVNFISFMEPLAKSPEHRQELIEBPETIPACEBILR	280
Db	224	DRISEGELIAMVFLLIYAGHETTIVNLIGNCVLSLGNPDOLAALRNDPSLLPAIESTLR	283
Qy	291	RESLVADG--RILTSDEYFNHGVOLKQDOLLPMOLSGLDERRNACPMHDFEOROKVSH	348
Db	284	YESPVANSTFRHTAEAVRFGDVVIREBELVWALGAANRNGGERFEDPDREDIRRETTGHV	343
Qy	349	TFFGSHLCLGOLARREIIVTLKWLITRIPDSIA--PCAQIOHKSIGVSGVALPL	404
Db	344	AFGIGIHFCVGAALARLEAQIANGRLLERPDLIRMAASPDLLMRVSEGLKLEKPL	401

RESULT	11			
CPXK_SACER		STANDARD:	PRT:	405 AA.
AC	CPXK_SACER			
DT	P332771;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	CYTCHROME P450 107B1 (EC 1.14.-.-) (P450CVIIB1).			
GN	CYP107B1.			
OS	Saccharopolyspora erythraea (Streptomyces erythraeus).			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Pseudonocardiaceae; Pseudonocardiaceae;			
OC	Saccharopolyspora.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 172-203 AND 302-321.			
RC	STRAIN=NRRL 23368.			
RX	MEDLINE: 92121109.			
RA	Andersen J.F., Hutchinson C.R.;			
RT	"Characterization of Saccharopolyspora erythraea cytochrome P-450			
RT	genes and enzymes, including 6-deoxyerythronolide B hydroxylase.";			
RL	J. Bacteriol. 174:725-735(1992)."			
CC	-I- FUNCTION: NOT KNOWN, PROBABLY INVOLVED IN THE CATABOLISM OF			
CC	OCTANE AND GUAIACOL. IT DISPLAYS A WEAK ACTIVITY IN THE O-			
CC	DEALKYLATION OF 7-ETHOXYCUMARIN.			
CC	-I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ,			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL: M83110; AAA26483.1; ..			
DR	PIR: B42606; B42606.			
DR	HSSP: Q00441; 10XA.			
DR	PFAM: PF00067; P450. 1.			
DR	PRINTS: PR00359; BP450.			
DR	PROSITE: PS00086; CYTOCHROME_P450. 1.			
KT	Oxidoreductase, Monooxygenase, Electron transport; Membrane; Heme.			
FT	BINDING 352 HEME (BY SIMILARITY)			
Q0	SEQUENCE 405 AA; 45238 MW; 71C93CEC1FD53FD CIRC64;			

```

RESULT 12
CPXY_BACSU          STANDARD:      PRT:      410 AA.
AC      008469;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-FEB-2000 (Rel. 39, Last annotation update)
DE      CYTOCHROME P450 (EC 1.14.-.-).
GN      CYP4 OR CYP107J1.
OS      Bacillus subtilis.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group;
OC      Bacillus/Staphylococcus group; Bacillus.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-168 / BGSJAI;
RX      MEDLINE; 97431495.
RA      Bellitsky B.R., Gustafsson M.C.U., Sonenshein A.L., von Wachenfeldt C.;
RT      "An lrp-like gene of Bacillus subtilis involved in branched-chain
RT      amino acid transport."
RL      J. Bacteriol. 179:5448-5457(1997).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-168;
RX      MEDLINE; 97453479.
RA      Sorokin A., Bolotin A., Purnelle B., Halbert H., Lauber J.,
RT      Duesterhoeft A., Ehrlich S.D.;
RT      "Sequence of the Bacillus subtilis genome region in the vicinity of
RT      the lev operon reveals two new extracytoplasmic function RNA
RT      polymerase sigma factors S19V and S19Z."
RL      Microbiology 143:2939-2943(1997).
CC      -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

	Query Match	15.3%	Score 333.5;	DB 1;	Length 405;
	Best Local Similarity	25.6%;	Pred. No 2.7e-18;		
	Matches 107; Conservative	74;	Mismatches 184;	Indels 53;	Gaps 10;
OY	18 VPEHLVED---FDMYNSLSAGVQEAAMAVLQESNVDPLVTWTRCGN-GHWIATRGQLIR	72			
	:   :	:	:	:	:
	:   :	:	:	:	:
Db	6 VPDLIAEADAFQAORHN-----RYARKREPVORI--RTVNGGDAMIITREYEDK	53			
	:   :	:	:	:	:
	:   :	:	:	:	:
Y73	EAYEDYRHFSSECEFIIPKEAGEAYDFITSM-----DPPGROGRF	112			
	:   :	:	:	:	:
	:   :	:	:	:	:

DR EMBL: Y11043; CAA71937.1; -;  
DR EMBL: U93876; AAB80898.1; -;  
DR EMBL: Z98117; CAB14615.1; -;  
DR HSSP: Q00441; 10XA.  
DR SUBTILIST: BG11929; CYPA.  
DR PFAM: PF00067; P450.1.  
DR PRINTS: PR00359; BP450.  
DR PROSITE: PS00086; CYTOCHROME\_P450.1.  
NM Oxidoreductase: Monooxygenase: Heme-



FT BINDING 359 359 HEME (BY SIMILARITY).  
SQ SEQUENCE 410 AA; 47384 MW; 035E98E58A4761AB CRC64;

Query Match 15.3%; Score 333; DB 1; Length 410;  
Best Local Similarity 27.2%; Pred. NO. 3e-18;

Matches 86; Conservative 69; Mismatches 139; Indels 22; Gaps 5;

OY 104 DPPQROFRALANOVGVYDKLENNRQELACSLIESLRQOGCNFEDYAEPPPIRIF 163  
||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 100 DPHNNRLFTLYQKAFTHRLMLQLEDKIQHTADSLDKVQKNKFMNLVDYAFPLPIVI 159  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 164 MLALGRLPEEDIPHLKYLTDMQTRPDGSMTEFAEAKAL-----YDYLPIIEORQ 213  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 160 SEMGPILEDPEKFRVMSQAL-----IDFSDAPERLOENDHLLGEFVEYELSLVRKKRR 213  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 214 KPGDAISIVANGVNGRPITSDAKRMCGLLVGGIDTVYNFLSFSMEFLAKSPENHOE 273  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 214 EPAGDLISALIQASESTQSTEEELYSIMILLIVAGHETVNLITNMVTALMCHHDOLEK 273  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 274 LIERPERIPACEELLRRESLV--ADGRILTSDEYEPHGVOLKKGDDILLPOMLSGLDERK 331  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 274 LRQOPDLNNSAIEELRHSFVELTTIRWTAEPFLHGOETKRKQVITITSLASANRDEKI 333  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 332 NACPMHVDPSROKVSHTTFCGSHLCUGOHARREIIVTLKEMLTRIPDFSIAPGAOTQH 391  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 334 FPNADIPDIEKKNNRHIAFGNHFCLGAQALRLAKAISTLARCPNIOL-KGEKKQM 392  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 392 K---SGIVSGVQALPL 404  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 393 KWKGNFLMLALELPI 408

RESULT 13

CPXE\_STRGO STANDARD; PRT; 405 AA.

AC P18326;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME P450-SU1 (EC 1.14.-.-) (P450-CVA1) (CYP105A1).  
GN CYP105A1 OR SUAC.  
OS Streptomyces griseolus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-31.  
RC STRAIN-ATCC 11796;  
RX MEDLINE; 90264332.  
RA Omer C.A., Lenstra R., Little P.J., Dean C., Tepperman J.M.,  
RA Leto K.J., Romesser J.A., O'Keefe D.P.;  
RT "Genes for two herbicide-inducible cytochromes P-450 from  
RT Streptomyces griseolus.";  
RL J. Bacteriol. 172:3335-3345(1990).  
CC -1- FUNCTION: METABOLISM OF A NUMBER OF SULFONYLUREA HERBICIDES.  
CC -1- INDUCTION: BY HERBICIDE.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M32238; AAA26823.1; -  
CC PIR; A35401; A35401.  
CC HSSP; P23295; 2ROM.  
CC PRAM; PF00067; P450; 1.  
CC PRINTS; PR00359; BP450.  
CC PROSITE; PS00086; CYTOCHROME\_P450; 1.  
CC Oxidoreductase; Monooxygenase; Electron transport; Heme.

FT INT\_MET 0 0  
FT BINDING 354 354 HEME (BY SIMILARITY).  
SQ SEQUENCE 405 AA; 44081 MW; 92AB36E064FD0B3E CRC64;

Query Match 15.1%; Score 330; DB 1; Length 405;  
Best Local Similarity 26.3%; Pred. NO. 5e-18;

Matches 104; Conservative 74; Mismatches 193; Indels 24; Gaps 11;

OY 31 PSNLSAGVQ--EAMAVIOESNVPLWTRCGH-WIATROQLREAVEDR----- 79  
||| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 16 PSNRSQVQLPDGYAQLKDPGCPHLRVTLYDGRQAMVVTKEHAKRKLGDRLSSNRTDD 75  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 80 HFSECFPIP--REAGEADYFIPMSDPPEQORFALANOVGVNPKLENNRQELACSL 137  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 76 NFPAISPREFAVRESPOAF-----IGDPEHGTGRRMTISEFTYKRIGKAMPVEEVNHG 131  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 138 LIESLRPOG-OCNFTEDYAEPPPIRIFMLAGLEEDIPHLKYLTDMQTRPDGSMTEFAE 196  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 132 FLDEMLAAGPTADLVSOFLPVPMSVICRLGVPYADHEFPQDASKRLVQSTDAQSALTA 191  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 197 KEALYDILPIIEOROKRGTDAI-SIVANGVNGRPITSDAKRMCGLLVGGIDTVYN 255  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 192 RNDLAGYLDGLITQFQTEPGAGLVADQDLANGE-IDREELISTAMLLIAGHETTAS 250  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 256 FLSPSEFLAKSPENHOELIERPERIPACEELLRRESLV--VAQGRILTSDEYEPHGVOLK 313  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 251 MTSLSVITLDPEDYALRADRSIVGAEVLELYLADIAGRAVTADEYEGHILIR 310  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 314 KGDOILLPOMLSGLDERKNACPMHVDPSROKVSHTTFCGSHLCUGOHARREIIVTLKE 373  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 311 AGEQIVVNSIANDRGVYVEDPDALDIHRSARHNHLAFGFGVHCGQLARLELEVLINA 370  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 374 WLTRIPDFSLA-PGAQIOHKSG-IVSGVQALPLVW 406  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 371 LMDRVPRLRAVPEQVLVRPQTITQGVNELPLVW 405

RESULT 14

YC56\_MYCTU STANDARD; PRT; 405 AA.

AC Q11062;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PUTATIVE CYTOCHROME P450 RV1256C (EC 1.14.-.-).  
GN RV1256C OR MYCY50.26.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RX MEDLINE; 98295987.  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogan A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandram M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrett B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

CC or send an email to license@ibb-sib.ch).
CC -----
DR EMBL; 277137; CAB00896.1; -.
DR HSSP; P33006; 1CPT.
DR TUBERCULIST; RV1256C; -.
DR PFAM; PF00067; P450; 2.
DR PROSITE; PS00068; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme-
FT BINDING 354 354 HEME (BY SIMILARITY)
SQ SEQUENCE 405 AA; 44580 MW; 72DEAB6C8688PA48 CRC64;

Query Match 14.7%; Score 319.5; DB 1; Length 405;
Best Local Similarity 29.5%; Pred. No. 3.2e-17;
Matches 102; Conservative 51; Mismatches 160; Indels 33; Gaps 9;

QY 62 HMIATRGQILREAYEDYRHFSSECPPIPREAGE-----AYDFITPSM-DPEPQRFALA 115
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 46 YVLSRHADVMSARDHDTFSS-AQGLIVNNGELEMIGLHDPVVMQDPVHTSEKLY 104
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

QY 116 NQVGMPPVVDKLENRIQELASLSLSPQGCNCTEDYAEFFPRIRIMLLAGLPEDIP 175
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 105 SRGFTPROVEVEPEYRFRFVERLEKLANANGGDIIVTELEFRPLSPVVAHYLGVEEDWT 164
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

QY 176 HLKYLTDQMPR---DGSMTFA-EAKELAYLYLPIIEORRQKPTDAIS--IVANGQVN 229
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 165 QFDGCTQAIIVANAVDAGTTGALDVGSMATFGLIRRKRTPEADDAISHLVAAVGAD 224
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

QY 230 GRPIYSDEAKRMCGILLVCGLDIVVNFPSFSMEFLAKSPENROELIERPERIPACEELL 289
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 225 GDTAGTSLIAFTFMVYGGNDVTYGMLGSMPLHRRPDQRRLDDPEGLPDAVEELL 284
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

QY 290 RRFSLVAD-GRILSDYFHHGVQLKKGDQILLPOMLSGLDERK-----NACPMHY 338
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 285 RLTSVQGLIARTFTTREDVYIGDTTIPAGRRVILLVGSANRDEQYGPDAELDVTGCRPNI 344
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

QY 339 DFRQKVSHTFFGSHSLGONHLARRILYTIKEMWLRIPIFSLA 384
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 345 -----LTFSGAHHCILGAAARMCQRAVALTELLARCPDEFA 381
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

RESULT 15
CPXL_PSESP STANDARD; PRT: 428 AA.
AC P33006;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, last annotation update)
DE CYTOCHROME P450-TERP (EC 1.14.-.-) (CYTOCHROME P450 108).
GN CYP108 OR TERPC.
OS Pseudomonas sp.
OC Bacteria; Proteobacteria.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 92332528.
RA Peterson J.A., Lu J.-Y., Geisselsoder J., Graham-Lorence S.,
   Carmoma C., Withey F., Iolence M.C.;
RT "Cytochrome P-450terp. Isolation and purification of the protein and
   cloning and sequencing of its operon.";
RT RT Biol. Chem. 267:14193-14203(1992).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE; 94166081.
RA Hasemann C.A., Ravichandran K.G., Peterson J.A., Delsenliffer J.;
RT "Crystal structure and refinement of cytochrome P450terp at 2.3-A
   resolution.";
RL J. Mol. Biol. 236:1169-1185(1994).
CC - FUNCTION: INVOLVED IN A ALPHA-TERPENEOL OXIDATION SYSTEM.
CC - SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```

CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; M91440; AAA25996.1; -.	
DR	PIR; S27653; S27653.	
DR	PIR; A42971; A42971.	
DR	PDB; 1CPT; 31-JAN-94.	
DR	PFAM; PF00067; P450; 1.	
DR	PRINTS; PR00359; BP450.	
DR	PROSITE; PS00086; CYTOCHROME_P450; 1.	
KW	Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;	
FT	3D-structure	
FT	BINDING	377
FT	TURN	3
FT	HELIX	9
FT	HELIX	16
FT	HELIX	19
FT	HELIX	22
FT	HELIX	24
FT	STRAND	40
FT	TURN	46
FT	STRAND	47
FT	STRAND	51
FT	HELIX	57
FT	TURN	67
FT	STRAND	69
FT	HELIX	70
FT	HELIX	82
FT	TURN	92
FT	HELIX	94
FT	HELIX	102
FT	TURN	107
FT	HELIX	108
FT	HELIX	109
FT	TURN	119
FT	HELIX	120
FT	HELIX	123
FT	TURN	145
FT	STRAND	146
FT	STRAND	150
FT	HELIX	152
FT	HELIX	153
FT	TURN	157
FT	HELIX	160
FT	HELIX	161
FT	TURN	171
FT	HELIX	171
FT	HELIX	174
FT	TURN	185
FT	HELIX	188
FT	TURN	208
FT	HELIX	230
FT	TURN	231
FT	HELIX	232
FT	HELIX	238
FT	STRAND	244
FT	STRAND	246
FT	STRAND	251
FT	HELIX	251
FT	HELIX	254
FT	HELIX	285
FT	HELIX	287
FT	HELIX	297
FT	HELIX	311
FT	STRAND	314
FT	STRAND	314
FT	STRAND	318
FT	STRAND	321
FT	STRAND	325
FT	STRAND	327
FT	TURN	328
FT	TURN	329
FT	STRAND	332
FT	TURN	330
FT	TURN	334
FT	STRAND	335
FT	STRAND	337
FT	HELIX	340
FT	HELIX	342
FT	HELIX	345
FT	TURN	346
FT	TURN	347
FT	TURN	349
FT	TURN	351
FT	TURN	355
FT	TURN	360
FT	TURN	361
FT	TURN	369
FT	TURN	370
FT	TURN	374
FT	TURN	375
FT	TURN	378
FT	HELIX	379
FT	HELIX	380
FT	HELIX	397
FT	STRAND	398
FT	STRAND	403
FT	STRAND	408
FT	STRAND	409
FT	STRAND	415
FT	STRAND	415
FT	STRAND	416
FT	STRAND	420
FT	STRAND	422
FT	STRAND	427
SQ	SEQUENCE	428 AA; 47922 MW; 298772460CC4E93F CRC64;

Query Match

14.68; Score 318; DB 1; Length 428;

Best Local Similarity 26.18; Pred. NO. 4.4e-17;  
Matches 100; Conservative 61; Mismatches 170; Indels 52; Gaps 9;

```

OY 63 WIAR-----GOL-----RAYEDYRHFSCEPFIIRREGEAVDFI 99
Db 52 WIAKHADWYMOIGKQPGELFSABESSELLYOONNEAF--MSISGCGHV-----IDSL 102
OY 100 PTSMDEPEQORFALANQVVMGYVYDKLENRIQELACLSSLIESLRP-OGOCNFTEDYAEFP 158
Db 103 -TSMDFPTHTAYRGELTLNMFQOPASIRKLEENIRRIAGASVO RLDPFGCEDFMTDCALY 161
OY 159 PIRIFMLLAGPREDIDHKLTYD-----QMTRPDGSMTFAEAKELYD 202
Db 162 PLAVVMALGVPEDDEFLMLKLTODFGVHEHPDEQAAVAPROSADAEARREHETIATFYD 221
OY 203 YLPIIEOBROKPTDASIVYANQVGRPTTSEAKRMGOLLVGLDVPVNFLEPFSME 262
Db 222 YFNGFTYDRKSCRPDDWMSLANSKLDGENTYIDKYNIAAYVALITAGHDITSSSGGAT 261
OY 263 FLAKSPEHROELLERPERIPACEBELLRFESLVAD-GRITSDYEFNGVOLLKGDQILP 321
Db 282 GLSRNPEQLAKSDPALIPRLVDEAVRMNAPVAFSMFKTALADTREVNGQINIKSDRIMLS 341
OY 322 OMSGLDERKNACPMHVDFSOKYSHTFPGHSHLCLGOHLARREITVTLKEWLTIRIPDF 361
Db 342 YPSANRDEEVFSNDEDEDITFRPNRHGFGWAGMCGLOHLAKLEMKIFFEELLPKLKSV 401
OY 382 SIAPGAQIOHKSGSIVGVOALP 404
Db 402 ELS-GPRLVATNFGVGGKNVPI 423

```

Search completed: October 4, 2000, 13:04:52  
Job time: 1683 sec

**THIS PAGE BLANK (USPTO)**